

Spring 5-2016

Validating Placement for Teaching and Learning

Jessica Ann Ganni

Follow this and additional works at: <http://knowledge.library.iup.edu/etd>

Recommended Citation

Ganni, Jessica Ann, "Validating Placement for Teaching and Learning" (2016). *Theses and Dissertations (All)*. 1413.
<http://knowledge.library.iup.edu/etd/1413>

This Dissertation is brought to you for free and open access by Knowledge Repository @ IUP. It has been accepted for inclusion in Theses and Dissertations (All) by an authorized administrator of Knowledge Repository @ IUP. For more information, please contact cclouser@iup.edu, sara.parme@iup.edu.

VALIDATING PLACEMENT FOR TEACHING AND LEARNING

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Jessica Ann Ganni

Indiana University of Pennsylvania

May 2016

© 2016 Jessica Ann Ganni

All Rights Reserved

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of English

We hereby approve the dissertation of

Jessica Ann Ganni

Candidate for the degree of Doctor of Philosophy

Michael M. Williamson, Ph.D.
Professor of English, Advisor

Ben Rafoth, Ed.D.
Distinguished University Professor

Lilia Savova, Ph.D.
Professor of English

ACCEPTED

Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: Validating Placement for Teaching and Learning

Author: Jessica Ann Ganni

Dissertation Chair: Dr. Michael M. Williamson

Dissertation Committee Members: Dr. Ben Rafoth
Dr. Lilia Savova

This study contributed to teaching and learning based writing assessment literature (Huot, 2002) by beginning the process of crafting a validity argument for writing placement at Cardinal College, a pseudonym for a small, private college in the suburbs of a mid-sized, Mid-Atlantic city, which uses Smith's (1992/2009, 1993) *expert reader model* to place students into one of two reading and writing integrated courses. The study specified the program's *interpretive argument* (Kane, 2006) and investigated one key assumption of that argument—that readers place students according to course-related criteria—by conducting interviews and think-aloud protocol sessions (Ericsson & Simon, 1984) with placement readers and course teachers while they read student work, by administering a placement adequacy survey to teachers, and by examining course documents and final grades.

The study found that Cardinal College's placement readers use teaching specific criteria to determine a student's position on a continuum of ability linking reading and writing, presenting a pedagogically useful view of students for the program. However, because of a change to the curriculum, with the loss of a mid-level course between the remaining courses, the study found that (a) the course designations and understandings from the lost course, particularly on reading, influence placement decisions more than the expert-reader model; (b) higher course teachers are not in agreement on the importance or

work of assessing students' reading; (c) yet, lower course teachers are strongly aligned and knowledgeable about troublesome reading issues.

It was the conclusion of this study that Cardinal College's writing program must formally address the lost course by having teachers share different views of the student population and strategies for addressing reading issues to redevelop its courses to meet all students' needs. Further, placement reading must be modified so that different course readers wait to place worrisome students and instead have conversations with each other, making the decision in concert using the expertise of all. In addition to the value to Cardinal College, the study has value to other writing programs at small liberal arts colleges, those with reading and writing integrated curriculum, and those undergoing curricular change or experiencing course misalignment.

ACKNOWLEDGMENTS

No one does any large and difficult task in isolation. Pursuing my doctorate degree has been a challenge, and for the most part it was not because of the work alone. Along the way, I have been helped and aided by so many people, primarily faculty, friends, and family, and I want to make sure I acknowledge their contributions here.

First, let me thank my participants, primarily the seven teachers at “Cardinal College” who agreed to spend the hours required for participation in this study: Adele, Daria, Hannah, Janet, Odysseus, Terence, and Theresa. I am grateful that you were willing to share your thinking on the writing of your students with me. I learned a great deal from being able to see through your eyes about this work we do, and I appreciate that gift. My deepest thanks, in particular, go to Adele and Theresa who also participated as placement readers, and thus gave up hours of precious summer break time to share their thinking on placement essays with me. While the students of this study are unnamed, I want to thank them as well. Without their willing hearts and the gift of their writing, I would understand far less than I do now about what it means to teach and assess writing.

Second, I want to thank the IUP students with whom I studied. It was a joy to learn in your presence and to share in your insights. I also want to thank the IUP faculty members who taught and mentored me, both in the classroom and through this research. I learned a great deal from your instruction and advice. In particular, I want to thank Sharon Deckert, who helped me through the many forms, procedures, and bureaucracy, as well as daunting and unique events in the process, like retirements and extensions. Also, thanks are due to Sue Welsh, who despite retiring before I could finish, was a

fabulous and enthusiastic reader of the first three chapters of this study, and to Lilia Savova, who was only added as a reader in the last months of work, but whose joy is always infectious. Thank you, Ben Rafoth. You were one of the first people to push me on the ever expanding nature of my research questions, and without your early critique, this study would have suffered. I am grateful for your careful and detailed review of my writing. And finally, thanks to my advisor, Mike Williamson. I've had the opportunity to carry on a prolonged conversation with you about writing assessment that started in your class and continued through this project. I can't express how much I've learned from that opportunity. Thank you for letting me find my way, for putting pressure and relieving pressure at just the right times, and thank you for the great respect you gave my work.

As I wrote this dissertation across years, a number of colleagues at several institutions where I teach cheered me on, inquired about my work, and listened to me talk about my research. I want to thank Bev Carmo, Mize Gardner, Sam Owens, and Iris Kyles, but especially my friend Patrice Shelly who occupied the same tutoring office as me, shared lunch time breaks, and always sent me back to work, even when I was tired and would rather do anything else instead. I also want to thank Sr. Rita Adams, Carol Moltz, Astrid Kersten, Sheila Mueller, Mark Dawson, Ed Stankowski, Marie Deem, Diana Forsythe, Josh Bellin, Janine Bayer, Nancy Wehrheim, Lisa Roulette, Michelle Maher, Katie Heiles, Jim Neutrelle, Mary Ellen Adams, and Barb Bencsics. Thank you for always asking how I was doing.

I am blessed to also call a number of fellow teachers and fellow writers my friends, and thanks are due for their support, whether given in person or through Facebook messages and comments. Thank you, Sr. Rita Yeasted, Joanna Crammond,

Penny Pence, Dana Olwan, Bri Lawson, Maggie Herb, Sean Armstrong, Marjie Stewart, Whitney Sarver, Dawn Fels, Brian Cope, Robin Cartwright Gallaher, Nicole Warwick, Zach Hull, Beth Ann Hull, Nick Sauer, Tina Perdue, Patricia Lietz, and Toni Bartone. More specific thanks are due to a few of these friends. Sr. Rita, thank you for all the conversations about teaching, all the encouraging words and prayers, and for your laugh. Joanna, thank you for writing assignments and discussing students with me, for problem solving the odd cases, and for time spent in lunch and friendship. Thanks especially to Penny Pence who, even though she didn't know me at the time, was visiting Pittsburgh and through a mutual friend had the opportunity to give me key advice at a critical moment in the development of this study. Thank you, Penny for saying that I just couldn't study all my research questions and subquestions on writing placement at once. Despite your advice, I still managed to squeeze in too much, but you tried.

I am thankful for the work of fellow teachers and writers in the field. While I do not know most of those people I cited in this study personally, as I was writing it felt like I did. I am particularly grateful for the writing and work of Brian Huot and Bill Smith. I never talked to either personally in the writing of this dissertation, but the conversations I carried on with both men in my mind are central to my thinking here. I hope I have done right by their work.

Next, I want to thank my family and friends. I am blessed to have so many friends in this world, many like family, and the love, laughter, and fun I share with them cannot be discounted. It is only if we have joy that we can keep going. Thank you Nancy Stevens, Sara Davis, Karen Davis, Beth & Scott Cahill, Amy Blyth-Whipkey, Elena Blyth, Morgan Crist, Katie (Ray) Campbell, Katelyn Ley, Kaetlyn Bowser, Jason Dicker,

Elliot Richards, and Donny Ohler. Thank you for all the lunches, chats, pop-culture discussions, pictures of dachshunds, and silly shenanigans. Most especially, to Sally Reffert who shares my love of board games and “junks.” At many points in this process, when a lot of life seemed like work, you offered simple fun, and I thank you for it.

I also have been blessed with a rowdy and boisterous family, both the family I was born into and the family I joined in marriage. I may be a glutton for punishment, but I find that opinionated, direct, and pushy people give me comfort and help me find my way. Thank you to the Ganni clan—Aunt Patty, Aunt Joann, Uncle Tom, Aunt Theresa, Aunt Mary-Ann, Uncle Jerry, Tara, Bob, Anthony, and Mia—but most especially to my “Dill,” my daddy-in-law, Big Larry Ganni. Thank you for the food, the inappropriate jokes, the silliness, and for giving me your son. To my family, in particular Kristy, Mom, Kayla, Haylee, Kylee, and Kellee, thank you for the love. And to my sister, Connie, thank you for coming back into my life. We’ve banged heads since they first put us in a room together, and our relationship has never been easy. Like two magnets with opposite charge, your soul challenges mine, but I am made better for the experience of a different view, and I am thankful we are trying to learn how to keep our energies from pushing us apart. Life would be so dull without you, and I am grateful for all the fun you and your girls bring into my life; I love you.

I have a few specific friends and family who require more detailed thanks, as without them, I would never have been able to do this work. Linda Jordan-Platt, I couldn't imagine my life without you in it. You were my teacher, and I learned from you about writing, linguistics, and literature. You mentored me when I began teaching; helping me learn how to see students in all their complexity and center on them and how

to reflect on my experiences to get better as a teacher. Through your example, I was able to navigate being a working class academic. I am also grateful that you let me move out of the "learner" role, as you treated me as an equal and learned from me. A number of times as I was working on this project, your thinking helped show me something I hadn't considered. Whenever I am working a problem where I know I need to see the idea in a different way conceptually to figure out where I'm stuck, I come to you.

Next, to April Julier, my Grapsie, I have always been blessed to have you as a cheerleader for my work. Most of my academic reading, thinking, and writing has been done in proximity to you—in classes, in the gazebo, in your living room, on the phone, and even through Skype. One of the most frustrating aspects of our recent life, for me at least, has been the change in proximity. Where I once could walk a block to eat Chinese food and write with you, I now would have to fly across the country. Since this "gig" we both have doesn't pay, my thin wallet keeps me from you, from being able to sit in the same space with you, and from hearing you talk, in your enthusiastic and forceful way, about the teaching of writing and the treasuring of what we learn from our students. I admire you so much for that, and despite this distance in space, thank you for bolstering me with constant posts and messages, cheering for me, supporting me, and giving me something to laugh at. Thank you for that last minute pep talk when I was floundering. I can't wait until I can save up some money to sit on your couch and talk teaching again.

To my "WC" crew—Anju Manandhar, Ashley Solomon, Cadwell Turnbull, Caitlin Bahr, Devin Patterson, Mike Hassett, and Therese Joseph—each of you has in some way walked with me through this process. Caitlin and Devin, while the miles have separated us from writing together lately, both of you helped me a great deal when this

project began. When I was knee-deep in books and articles, trying to find my way through a proposal for this study, I remember the time I spent sitting next to both of you and writing. Mike, a good deal of writing in the middle of this project happened right when you needed a place to crash and you became the first guest in “the guestroom.” You and I spent every day not only working together, but commuting and living together. We wrote, played music, drank coffee, and ate “bacon gloriousness” omelets. We talked endlessly about everything and conversations blurred into all-day discussions of whatever as we spent the day laughing through our work. I will forever treasure that time.

Anju and Cadwell, it is true that whenever we were together in the same location, work on our individual projects has taken place. But more importantly, you two provided and still provide nourishment to my soul. Whenever you are around, I feel loved and understood, safe to admit my fears and cry, free to play and be silly, and I get to have the kind of conversations that teach me something about this quest we are all on in life and how to live with principle and purpose. Ashley and Therese, I saved you both for last because I believe you guys carried me through the much more difficult part of this journey. When I was almost finished, but not quite, I was able to depend on you guys to show up without fail every week, eat pizza and candy with me, and put in the hours of work sitting across the table from me. You frequently had to deal with my whining, my worry, and my craziness. I couldn’t have had a better team on those difficult nights than “shit-can” Therese and “sassy” Ashley, using all means necessary to set me right again.

To my niece and nephew, Autumn and Justin Barszczowski. I have had the privilege to not only watch you both grow from babies into adults, but to have had a hand in the process, and I couldn’t be prouder of you two. As I was completing this project, I

was constantly bolstered by your interest, your pushing, your teasing, and your love. I think most people would be embarrassed to let children see them struggle or feel unsure, but I've always tried to show you that learning is hard work and you have to keep pushing to better yourself, even when it is difficult. Thank you both for handing me back that lesson when I was losing faith and running low. Thank you especially, Autumn, for showing up to do your writing while I was doing mine. Thank you both for loving me. Thank you for being the only two people on the planet who think I'm cooler than your dad. Thank you for making me your "Iss."

To Nessa, the first time I imagined writing this section of my acknowledgements page, I imagined what I would say to you *and* Mike. How I would thank you both for opening your house for all those writing days. The way you both listened to me cry when it was hardest—when I doubted myself and didn't feel I was good enough, when I spoke my deepest fears aloud and found comfort in an unending supply of tissues, Gram Parson's music, Poppet puppy kisses, Mike's fried rice, and his chocolate and pistachio cake too. I never imagined he wouldn't be with us on the other end of this project. I never thought we'd be here without him. Losing him is still so fresh and new, I am not exactly sure what to say now. Everything I might say seems inadequate, except maybe that I love you. I want you to know that I could never have done this without you and Mike, and without your support. You both provided me with strength to keep going when there seemed no way to balance everything life was throwing at me. Hopefully I can provide the same unending support and love to you now, and we will both keep walking forward.

To my brother, Josh Barszczowski and to my daddy, Richard Barszczowski. I am a lucky, lucky person to have you both. Josh, it is a privilege to get to go through life with

someone who understands me on a microscopic level. You and I shared the same upbringing and culture, but we also share the same principles, chief among them that life is all about learning and teaching. I have done both learning and teaching officially in a classroom most of my life, but you do it by being you. I know you feel that you weren't enough help at times, that you wish you could have done more. But I can't stress enough that you taught me years ago to go through life doing the things that make me feel at home in my own skin. And, you continually provided, through your very presence, a reminder that I will be okay if I have that.

Daddy, I think when most people look at you, they see a pain in the butt. They see someone who pushes, who has his agenda, who is opinionated and difficult to get along with. You are not easy, that's for sure, but I have always been able to see through all that to your heart. I know you push when you are worried for me. I know your opinion, while not always one I share, is offered because you want me to factor it in. When we clash, it is fierce and heated, but it is because we have passion and love for the people we care about in a world fraught with the possibility of peril. Thank you for everything, big and small, you did through this process like buying me lunch after lunch, driving me back and forth to IUP when I was under slept, taking me to the woods to write, showing up and cooking meals while I was writing, loaning me money, paying attention to things around the house when I couldn't, and not being insulted when I was only half listening. But most importantly, thank you for making me a priority. I love you.

To my love, Larry Ganni, there will never be enough thanks for all the things you did to help me in this process. Thanks for all the laundry, chores, meals and errands, and for essentially taking on all the daily tasks of our lives that were once shared and doing

them on your own for years without much help from me. Thanks for the back rubs, the kisses, and the wiping of tears. Thank you for making me laugh. Thanks for spending the last month of this project making up reasons to sit across the table from me as I wrote so that I would feel less alone. But thank you for so much more—for what is larger than this one project, for what is the whole of our lives together. Thank you for being you, and thank you for letting me have a way to be me. Thank you for accepting me how I am and loving me, but pushing back when I'm too much. Thank you for listening and for not. Thank you for giving me all of your attention, and then for being tired enough to say you were done for the day. Thank you for being willing to learn and change, and for asking me to do the same. I know that you tend to think you go unnoticed or unheard, but it's not true. When you act and speak from your heart, it is impossible not to notice you. You shine. And I'm lucky because with me, you are all heart. Thank you for your heart. It has been the best gift of my life.

On a final note, I want to dedicate this dissertation to the memory of all those I've lost during this journey. I'm not exactly sure how I survived and persisted through each of your passing, but I like to think it is because of the blessings that each of you gave me through your life and because of those I believe each of you are still giving me now. I hold you in my memory and have learned from the lives and example of each of you. Learning never dies since we pass it on as we share our knowledge with the world. As I share what each of you has taught me, your spirit lives on. I love you Grama, Rube, Tori, Uncle Mac, and Mike.

TABLE OF CONTENTS

Chapter	Page
ONE	INTRODUCTION.....1
	Overview of the Chapter2
	A Division Between Teaching and Assessment.....4
	Reconnecting Assessment and Teaching: Stepping Forward All Together7
	Concerns of Adequacy at Cardinal College11
	The Focus and Research Question15
	Data Collected and Analyzed17
	The Findings and Recommendations of This Study18
	The Value of Validation Research20
	An Overview of Upcoming Chapters24
	Conclusion.....29
TWO	REVIEW OF LITERATURE.....31
	Two Sides of Writing Assessment35
	Changes in the Definition of Validity (and Its Relationship to Reliability).....37
	Parallels Between Validity and English Teacher Concerns48
	Centering on Validity62
	William L. Smith: Validity, Expertise, and Context69
	The Relevance of Validity, Expertise, and Context in Placement Today105
	Conclusion.....113
THREE	METHODS AND PROCEDURES116
	Focus of the Study and Research Question119
	Overview of Validation120
	Outlining Cardinal College’s Context and Interpretive Argument.....124
	Key Assumptions of the Interpretive Argument: Selecting the Study Focus141
	Rationale for Study Design143
	Steps of Data Collection.....175
	Steps of Data Analysis200
	Strengths and Limitations of Data Collection and Analysis222
	Conclusion.....230

Chapter	Page
FOUR DATA ANALYSIS RESULTS AND FINDINGS	232
Chapter Four: Brief Overview.....	233
What Are Expert-Readers' Placement Criteria?	237
Are Expert-Readers' Placement Criteria Teaching Specific?	327
Are Expert-Readers' Placement Criteria Linked to Course Criteria?	346
Are Placement Concerns and Decisions Linked to Course Concerns and Decisions?.....	403
Conclusion.....	417
FIVE CONCLUSIONS AND RECOMMENDATIONS.....	420
A Brief History of a Divided Field for Writing Assessment.....	420
Reconnecting Assessment to Teaching and Learning: Validation	421
Why Was Validation Useful to Cardinal College?	424
Study Focus and Research Question	425
Study Design and Participants.....	427
Summary of the Study Findings	428
Conclusions and Discussion	430
Recommendations	433
Future Studies.....	441
Value of the Study for Cardinal College	445
Value of the Study for the Field	446
REFERENCES	451
APPENDICES	469
Appendix A - Writing Placement Prompt	469
Appendix B - Email to Placement Readers	472
Appendix C - Informed Consent Form: Placement Readers	473
Appendix D - Informed Consent Form: Placement Students	475
Appendix E - Placement Reader Initial Interview Questions.....	477
Appendix F - Think-Aloud Directions and Practice for Placement Readers	478
Appendix G - Email to Course Teachers.....	480
Appendix H - Informed Consent Form: Course Teachers	481
Appendix I - Informed Consent Form: Students	483
Appendix J - Course Teacher Initial Interview Questions	485
Appendix K - Think-Aloud Directions and Practice for Course Teachers.....	486
Appendix L - Survey of Placement Adequacy	488

Chapter	Page
Appendix M - EOTR Course Results	489
Appendix N - Summary of All Participating Students’ Demographic Information	490
Appendix O - Summary of Decisions on Participating Students in Original Placement Sessions	493
Appendix P - Summary of Decisions on Participating Students in Redo Placement Sessions	494
Appendix Q - List of All Unproblematic Placement Students	497
Appendix R - List of All Problematic Placement Students	499
Appendix S - List of All Problematic Placement Students With Their Problematic Designation(s)	501
Appendix T - List of Course Students’ Survey of Adequacy Ranking in Diagnostic Sessions	503

LIST OF TABLES

Table	Page
1 Think-Aloud Coding Themes	208
2 Reading and Writing Criteria.....	240
3 Total Number of Students Discussed for Each Placement Criterion.....	247
4 Total Number of Unproblematic Course 11 Students Discussed by Theresa for Each Placement Criterion	254
5 Total Number of Unproblematic Course 10 Students Discussed by Adele for Each Placement Criterion	264
6 Total Number of Unproblematic Course 10 Students Discussed by Theresa for Each Placement Criterion	274
7 Total Number of Problematic Course 11 Students Discussed by Theresa for Each Placement Criterion	280
8 Total Number of Problematic Course 11 Students Discussed by Adele for Each Placement Criterion	281
9 Total Number of Problematic Course 10 Students Discussed by Adele for Each Placement Criterion	300
10 Total Number of Problematic Course 10 Students Discussed by Theresa for Each Placement Criterion	301
11 Total Number of Course Students Discussed for Each Course Criterion	354
12 Total Number of Course 11 Students Discussed by Theresa for Each Course Criterion.....	356
13 Total Number of Course 11 Students Discussed by Odysseus for Each Course Criterion.....	360
14 Total Number of Course 11 Students Discussed by Terence for Each Course Criterion.....	366
15 Total Number of Course 11 Students Discussed by Janet for Each Course Criterion.....	374

Table	Page
16 Total Number of Course 10 Students Discussed by Adele for Each Course Criterion.....	380
17 Total Number of Course 10 Students Discussed by Daria for Each Course Criterion.....	385
18 Total Number of Course 10 Students Discussed by Hannah for Each Course Criterion.....	390
19 Total Number of Course 10 Students Discussed by Theresa for Each Course Criterion.....	394

LIST OF FIGURES

Figure	Page
1 Comparison of readers' Course 10 criteria	275
2 Comparison of Theresa's Course 10 and Course 11 criteria	276
3 Comparison of readers' problematic Course 11 criteria	283
4 Comparison of readers' problematic Course 10 criteria	303
5 Reading issues continuum.....	323
6 Assignment continuum	324
7 Essay and paragraphing continuum	324
8 Work with text continuum	325

CHAPTER ONE

INTRODUCTION

With a college degree increasingly becoming essential in today's world, it is not uncommon to see articles and editorials in the popular news that discuss issues of higher education such as college readiness, remediation, placement, and admissions. In recent years, news articles question the effectiveness of writing and math placement tests and procedures as they report on the findings of larger research studies. For example, Lewin (2012) reported on studies by Belfield and Crosta (2012) and Scott-Clayton (2012)—both out of the Community College Research Center of the Teacher's College at Columbia University—in order to ask if students, particularly those placed into developmental courses, need to be there. These articles raise many questions: Do students need to take developmental courses? How do college administrators and teachers decide that a student requires additional courses? What adverse effects do such courses have on retention? Do pre-college level courses make a difference in students' success in college?

While I find some of the premises and assumptions of Lewin's (2012) article and its cited studies problematic—a discussion I will return to shortly—I cannot deny that the basic question at the heart of the article is vitally important. How do educators know that we are making the best decisions for students? In terms of placement, when a school or program makes a decision about a student's writing placement, how does the school know that it is *adequately* (Smith 1992/2009, 1993) placing the student in the course that is the best fit for the student's learning needs? How does the school ensure that it is not, instead, negatively affecting students by costing them, and the educational system, unnecessary expense or difficulty? All of these questions are questions asking for some

assurance—some validation—that adequate decisions are being made. But historically, determining what is adequate or valid in terms of writing assessment has depended on who is answering the question.

While I will detail more precisely the specific research question, approach, and research data of the current study later in this chapter, it is important to know now that this study addresses just such concerns and issues of placement adequacy (Smith 1992/2009, 1993) as it investigates the placement process at a small, private college in the suburbs of a mid-sized, Mid-Atlantic city that, for the purposes of this study, will be referred to as Cardinal College (a pseudonym). Let me begin by first offering an overview of the chapter.

Overview of the Chapter

Before I go on to detail the specifics of the current study, some background information is necessary to put the important issues raised above about writing placement adequacy and validation into the larger historical context and the context of this study. Therefore, in the next section, I offer important historical background that will help contextualize the current study. First, I briefly detail the historical divide that existed between the field of testing (those who studied and developed tests and assessments) and the content specialists responsible for teaching writing. I want to make it clear both that these two separate groups were guided by their own theories and concerns and that the concerns of writing teachers were not given much voice in the early development of the field of writing assessment. The result was an unequal power division in writing assessment that favored the big business of testing and that created a long standing relationship between government and testing agencies. Second, I posit that it is important

to understand the current theory of writing assessment that has developed in the scholarship of writing teachers since it reestablishes the link between teaching and assessment in ways that emphasizes the contextualized work of teaching and learning. Finally, I discuss how the view and understanding of writing assessment that populates the news and receives national attention today is not typically that of writing teachers; instead, it is a view that has largely arisen out of a cultural familiarity with large-scale, national testing which brings with it a mistrust of teachers and an oversimplification of teaching and learning.

After discussing the historical division of writing assessment, I go on, in a subsequent section, to discuss what it will take to reconnect assessment and teaching. Of primary importance for writing teachers is how to address the concerns of our own field while also addressing the concerns of other stakeholders interested in the work of writing assessment (testing specialists, administrators, parents, students, etc.). I suggest that the argument of Huot (2002) that the writing assessment concerns of teachers can be supported and argued for through the understanding and use of educational measurement theories of validation is a powerful step forward. These theories, as Huot has shown, no longer emphasize the psychometric tradition that led to a context-free, large-scale, abstract, one-size-fits-all view of assessment which, in early writing assessment history, firmly discounted the local goals, particular contexts, and writing theories forefronted by writing teachers.

With this background information and context established, I turn to Cardinal College and this study in particular. First, I relate the historical reading of writing assessment literature I've sketched to the context of the particular site studied here,

establishing the concerns of adequacy at Cardinal College that necessitate study, specifically checking the assertion that placement readers at the College use course related criteria to make placement decisions and investigating the loss of an old course in the writing program.

Then, I specify the focus and research question of this study, the data collected and analyzed, and the findings of this study, including recommendations made for adjustments to Cardinal College's placement process. After, I discuss the value of validation research for Cardinal College specifically, and I follow that up with a discussion of the value of this study more broadly for the field of composition, and specifically the work of writing assessment. With the goals, findings, and value of the study established, I close this chapter with an overview of the chapters that follow.

A Division Between Teaching and Assessment

The questions of adequacy I outlined at the start of the chapter do not appear only in national sources like newspapers or the policy studies of educational institutes; they are deeply important to the reflective, professional educator. In fact, they are questions that have long been asked by writing teachers about the work of testing; it seemed to writing teachers that early assessments that did not actually collect students' writing could not adequately evaluate those students' writing abilities. Yet, since the early work of writing assessment had been developed primarily by researchers working in the field of testing or psychometrics, and only later contributed to by writing teachers, it is no surprise that the two different communities and their conflicting concerns, focuses, values, and conceptions of writing (Williamson, 1993) have been at odds.

Across the history of writing assessment, a seeming war (White, 1993) has been waged as teachers fought for ways of assessing writing that adequately addressed the reality of writing and the needs of teachers and learners. However, looking at the early history of writing assessment (1870s to 1960s), it is clear that teachers' concerns went unheeded since theories of writing were seldom reflected in early assessments. Instead, the commitment of many test specialists and psychometricians was to considerations of cost, efficiency, and objectivity (which, at the time, was understood as requiring decontextualized assessments) since these considerations were necessary to meet the growing demand for tests that could be administered to large numbers of people (Elliot, 2005). Testing agencies like the College Board and ETS were developed, and they focused on creating large-scale, standardized assessments with the goal of assessing and measuring ability regardless of whether it was the ability of U.S. Army personnel for government decisions of recruit placement or the abilities of an increasing population of college students for educative purposes (Elliot, 2005).

The reality and emphasis on large-scale testing concerns—coupled with the cultural belief that a standardized science of testing could be wielded to fairly and systematically determine the ability and merit of all—made it difficult for test creators to hear the writing theory and teaching-based concerns that were continually raised by English teachers (Elliot, 2005). Such concerns questioned if fair and instructionally significant decisions could be made without a sense of the specific details of writing theory and without regard to the particular contexts of teaching. Yet, even after writing teachers like White (1973) became involved in writing assessment, the work typically still stressed method and practice over theory and still relied heavily on testing concepts

(like reliability) that were part of a problematic tradition that was not congruent with the teaching of writing (Huot, 1996/2009).

Seeking to fill the gap, and to argue against past assessment practices that did not align theoretically with the teaching and theory of writing, scholars like Huot (1996/2009, 2002) and Huot and O'Neill (2009)—as well as the Conference on College Composition and Communication (CCCC) in its influential policy document *Writing Assessment: A Position Statement* (2009)—have argued for and begun developing theory to guide ongoing writing assessment work. One major focus in this work has been on reestablishing the primary purpose of assessments: to support teaching and learning goals by informing and enriching teaching and learning. Current theory claims that writing assessment should, among other things, be site-based and locally-controlled (Huot, 1996/2009) with pedagogical concerns playing a key role in how writing assessments are developed and refined (CCCC, 2009).

These kinds of locally developed assessments stand in direct contrast to large-scale, standardized assessments. Yet, it is large-scale assessments that are usually the focus of those outside the field of composition. The large-scale, national assessments of early testing history have again become the push of modern-day policy makers, politicians, and corporations who seek to institute standardized tests and curriculums throughout school—even into higher education—with little concern for local teaching contexts, the reality of student populations, or the theories of content areas, and with little understanding of how one goes about learning, teaching, or assessing more than just low-level recall and regurgitation.

Returning to the early discussion of Lewin (2012) and the cited studies of Belfield and Crosta (2012) and Scott-Clayton (2012), it is interesting to point out that while the viewpoints of Lewin, Belfield and Crosta, and Scott-Clayton are critical of standardized placement tests—as many writing teachers would also be—for the most part, the critique of these pieces does not question the assumption that a one-size fits all method of assessment is desirable. Instead, the assumption is that the right standardized placement measure has not yet been found. So while the studied placement measures are criticized as potentially inadequate for determining college success, this judgment is passed without considering the use of this test within the local teaching and learning goals of the studied colleges. What constructs of writing do these tests operate within, and what is the relationship of the employed test to the local curriculum? The problem with the existing placement systems used in these cited studies, I would argue, arises largely because these systems ignore questions of local context, teaching and learning, and the fit of the test to the situation. In short, these placements do not start from what the field of composition would consider sound writing assessment theory. Yet, the critiques presented by the authors ignore these matters in favor of a faith in the search for a standardized, one-size-fits-all approach that adequately places students.

Reconnecting Assessment and Teaching: Stepping Forward All Together

So what should the writing teacher—the writing program do? On one hand, we have our emerging theoretical understanding of what it takes to conduct meaningful and instructionally informative writing assessment—which tells us to focus on our local contexts and goals and to develop assessments that will support and enrich our work. On the other hand, there is a growing misunderstanding of teaching and mistrust of teachers

that keeps the voices of teachers unheard as policies, assessments, and decisions that we know will likely be damaging instead of uplifting to the education of our students are made. How do we right these opposing forces?

Clearly we must develop our own local assessments, as Huot (2002) and others have advised, because doing so makes the most sense for what we know about the role of context in the theories of reading and writing and in the tasks of teaching and learning. But how can we develop these assessments when the work of teaching—and the role of assessment in teaching—is misunderstood or mistrusted? As Huot (2002) sought to facilitate the development of local assessments, he argued that newer theories of validity in educational measurement would actually provide writing teachers a structure for validating and making a case for their local writing assessments to various stakeholders. They would be useful, he explained, because rather than presenting obstacles for writing teachers, as many prior versions of such theories had, newer theories had begun to recognize the “social construction of knowledge” (Huot, 1996/2009, p. 161) which aligned testing theory with writing theory and provided writing teachers with a way to make testing theory useful. I believe his suggestion provides one possible step forward.

While terms like validity and reliability come out of a psychometric tradition that once stressed large-scale, objective, one-size-fits-all tests over context specific assessment, such terms have not remained static. As the field of educational measurement has developed, the theories behind these terms have been revised. Newer theories of validity stress the need for ongoing study into an assessment process with the goal of crafting an argument for the adequacy and appropriateness of the use of an assessment, including thought given to the alternative arguments, the decisions to be made using that

assessment, and the consequences of its use (Cronbach 1988; Kane, 2006; Messick, 1989a, 1989b). It is these newer theories that Huot (2002) argued could be useful for investigating and explaining our local writing assessments since they require that the particular use of an assessment in a particular context is what is being investigated in validation. Additionally, these newer theories make ongoing validation part of the responsibility of test users and creators who must now consider alternative explanations and the challenges of their most critical stakeholders (Cronbach, 1971).

The responsibility inherent in validation means that while we may feel confident that we know what we are doing, we must investigate if our assessments are working sufficiently, and we must address the views and concerns of stakeholders (students, administrators, and the larger community) as we do so. Returning to where I began this chapter—with a focus on writing placement—the basic goal of placement is to fit a student into a course that will best serve their learning needs. That goal means that those designing and working within writing programs must investigate how they determine the needs of their students relative to the various courses they have designed. Yet, it also means that a system of placing students operates inside the larger work of a writing program where the program must consider, among other things, if their student population is changing and therefore exhibiting different needs, or if they have sufficiently articulated the goals and work of the courses in the program.

The theory or understanding from which a writing program operates—both in terms of their courses and their placement process—becomes an argument they make for the adequacy of their placement decisions. And it is this very argument that must be investigated in validation research. This way of understanding validation is supported in

the most recent explanations of testing standards and validity theory: Kane (2006) explained that to create a validity argument, any assessment or testing program must make a case for its process by first specifying *the interpretive argument* that it operates from, and then studying the assumptions behind that argument through research and the testing of rival arguments which lend support to the *validity argument* the program can go on to craft (p. 22).

Validation research is already being carried out by specific writing programs interested in investigating their placement systems, and just as Huot (2002) suggested, the use of educational measurement theories has lent such studies an air of transparency and frankness. For instance, recent studies like Gere, Aull, Green, and Porter (2010) and Gere, Aull, Perales Escudero, Lancaster, and Lei (2013) have focused on validating a local system of directed self-placement, and they have found interesting disconnects between the view of writing expressed in their placement documents and the one expressed in their course work (Gere et al., 2010). These findings have helped them refine their self-placement questionnaire and procedures going forward (Gere et al., 2013). Additionally, Elliot, Deess, Rudiny, and Kamal (2012) have started an investigation into the use of the standardized ACCUPLACER writing test at their particular site. Such a focus may, at first glance, seem unrelated to my point about local assessments, but even though Elliot et al. were not using a locally-created test, their validation work investigated the fit of the test to their local goals, courses, and context. Further, their work fit with validation theory in that the use of any test (even a test purchased from a testing company) must be validated in current test theory because the test itself can no longer be considered valid or invalid alone without some investigation

into the particular context of its use (Cronbach, 1971). Moreover, Elliot et al.'s report on their validation procedures detailed the difficulty with investigating the use of a test when testing companies are not willing to share test design information, and in doing so, their piece can be seen as presenting some of the clearest support for locally developed tests.

Concerns of Adequacy at Cardinal College

Now that a historical context has been established through a brief discussion of key literature on writing assessment, an understanding of current writing assessment theory, and an explanation of the work of validation, I discuss how the larger historical context of the field of writing assessment relates to the site studied here, Cardinal College. In short, this study sought to make use of Huot's (2002) suggestion that newer validity theories are useful for responsibly checking the assumptions of a writing program and addressing the concerns of its interested stakeholders by beginning validation work on the placement process at Cardinal College, and in that way adding to recent validation literature.

Cardinal College had an emerging set of concerns not unlike those sketched in the history of writing assessment above that could and should be addressed through ongoing validation research. First, we had been operating on the assumption (and had made the argument to administrators) that we use course criteria to make course relevant placement decisions, and we wanted to explore that assumption, as it was important to have data to support the assertions of our interpretive argument and to address the push from the administration for the implementation of an indirect test that would have faster score turnaround. Second, several years ago (around 2008), a four-credit course that fell between the two remaining courses in the program was eliminated by the administration

to regulate credit hours, and writing faculty had expressed concern about the loss of that course and the potential for students to not fit as neatly into the remaining course options. Let me sketch these two concerns a bit more to show why it was felt validation research would allow Cardinal College to investigate these concerns and learn valuable information for teaching and learning.

Making Placement Decisions Using Course Relevant Information

Some years ago (around 1998), Cardinal College adopted Smith's (1992/2009, 1993) *expert-reading model* to make placement decisions. Since the College uses this model to place students in reading and writing integrated courses on the basis of one, impromptu essay, some additional background on Smith's work is in order. In two published reports on decades of study into the placement process at the University of Pittsburgh, Smith (1992/2009, 1993) was able to examine—among other things—the adequacy and reliability of his placement readers' decisions. Through this investigation, Smith determined that the most adequate and reliable placement decisions were made by readers who focused on the course they had most recently taught and decided if a student belonged in that course or did not. When these readers rejected students as too low or too high for their course, additional readers, who were expert in the adjacent courses, would go on to read the student's essay and again determine if the student belonged in their course or not.

Since the adoption of Smith's (1992/2009, 1993) model, the College assumes that because it uses placement readers who have most recently taught particular courses, it is making course-related placement decisions. Cardinal College has asserted that expert placement readers tap their teaching knowledge and experience to make more adequate

placement decisions since those decisions are based on a locally-developed and direct writing assessment that is read using criteria that align with the writing program courses. In contrast, the placement decisions are not made using general observations on students' abilities that are disconnected from the courses, like those typically gained from indirect assessments and national placement options such as ACCUPLACER or COMPASS, which are standardized and therefore decontextualized assessments.

Drawing on Smith's (1992/2009, 1993) model, Cardinal College's writing program has pushed back against the adoption of a nationally-developed placement test, choosing instead to assert its locally-developed test and the expertise of its teacher-readers. Yet, Smith's model was based on formal study into his placement process, and such work has never been done at Cardinal College. Therefore, validation research can help Cardinal College investigate its assumptions and either substantiate its claim or better perfect its placement process to meet its goals.

Curricular Change From the Loss of a Middle Course

As I said earlier, Cardinal College recently lost a four-credit course that fell between the two remaining courses in the program when, due to space constraints on campus, the administration eliminated courses with odd credit numbers to regulate credit hours and fit courses into standardized time slots for scheduling ease. It is important to have some background on Cardinal College's courses to better understand the way the remaining courses operate and how the old configuration used to fit together.

Cardinal College has two existing first-semester writing course options—Course 10 and Course 11—that students can be placed into, and it used to have a third option that fell between the two existing courses, Course 11A. These courses were designed with the

College's large international, underprepared, and first-generation college student population in mind as the writing program overall stresses the development of academic literacy skills in both reading and writing to give students an "opportunity to learn" in a discourse that is potentially quite unfamiliar to them (Gee, 2003, p. 1).

Course 11, a three-credit college-level writing course, introduces beginning academic writers to the interpretive reading of complex academic texts—including the process of developing a reading and revising that reading through their writing (Bartholomae & Petrosky, 2004). And it develops students' understanding and abilities with academic writing forms and conventions. All students at the College must complete Course 11 (as well as a more research-focused course, Course 12), but since students entering the composition sequence have different abilities, and some students may not be ready for Course 11, the program has an additional placement choice: Course 10.

Course 10, a six credit class (where all credits count toward graduation and financial aid), prepares students for the kind of reading and writing found in Course 11 by immersing them in intensive reading and writing activities (Bartholomae & Petrosky, 1986). Students are given assignments nearly identical to those in Course 11, but they are given more instruction, scaffolding, and modeling throughout the reading and writing processes on such things as reading strategies, note-taking during reading, developing a position in relation to a text, essay and paragraph structure, the presentation of quotations, peer feedback, self-assessment, and identifying and controlling error. Additionally, with the endorsement of the instructor, a student who does well in Course 10 could have the course count for their Course 11 requirement.

The lost course, Course 11A, was identical to Course 11, the regular college writing option, in objectives and content and counted identically in completion of the College's first writing requirement; the difference was that the course was taught with an additional credit and hour per week for specialized time spent on developing close reading skills and working on error control for students who were in need of some additional scaffolding but not the six credit, Course 10 option. With the loss of this course that fell between the two other course options, the writing teachers have been concerned about the potential for students to not fit as neatly into the remaining course options.

The courses were crafted to be developmentally connected and imagined for Cardinal College's particular student population, which displays a diverse range of academic reading and writing abilities; therefore, the loss of the middle course has disturbed the intended boundaries of the courses, and it is not clear what effect the change has had. What decisions are best for old Course 11A students? Which Course 11A students would fare better in Course 10 and which in Course 11? And, how should the curriculum adjust to account for changing course-student populations? Again, the answers to these questions had never been determined through formal study at Cardinal College. Yet, with its emphasis on evidence gathering and investigation, validation research can help Cardinal College can study this concern.

The Focus and Research Question

Given Cardinal College's adequacy issues, this study was proposed to begin the process of crafting a validity argument for Cardinal College's placement by specifying the program's interpretive argument and investigating a key assumption of the writing

program's argument—that placement readers place students according to course content criteria.

Specifically, the study asked and answered the following research question:

- Do expert readers at Cardinal College use curriculum/course related and teaching specific criteria to offer a pedagogically useful and purposeful, view of students' placement essays? Are expert readers' criteria (and by extension their decisions) linked to course criteria?

When broken down, this research question divided into the following four subquestions that could be answered more directly with data:

1. What are expert-readers' placement criteria?
2. Are expert-readers' placement criteria teaching specific?
3. Are expert-readers' placement criteria linked to course criteria?
4. Are expert-readers' placement concerns and decisions linked to course concerns and decisions?

Before describing how the research question and subquestions were investigated, it is important to first stress that validation argumentation is complex work, and no one study can investigate all the assumptions and inferences of a program's processes.

Therefore, future studies must continue the work of validation at the College—pursuing other key assumptions, lines of inquiry, and data sources. For example, Cardinal College also assumes, as part of their interpretive argument, that the essay prompt used during placement presents a course-related task fruitful for revealing these criteria, and future studies would have to gather data on this and other assumptions as well (although the

data gathered can be reanalyzed to provide some initial insights on the placement prompt).

Data Collected and Analyzed

The research question and subquestions of this study were investigated in a number of ways through multiple data sets. By conducting interviews (Denzin & Lincoln, 2008) and think-aloud protocol sessions (Ericsson & Simon, 1984) with two placement readers while they read placements and with seven course teachers while they read student diagnostic essays and course work, the assumption that Cardinal College uses course-related criteria in placement was examined. Through analysis of these interviews and the think-aloud sessions where readers and teachers discussed a total of 119 participating students' work (49.17% of the total number of students enrolled in Course 10 and Course 11 for the semester of study), the study uncovered common and recurring criteria that emerged in the placement of each of the two courses at the College—Course 10 and Course 11.

Using these criteria, a relationship was explored between placement and teachers' course criteria that offered one data set relevant to the research question. Additionally, the research question was also examined through teachers' impressions and course grades. As the course teachers completed a survey tallying the number of students overall who they felt did or did not fit well into their course, as they detailed their impressions on the adequacy of placement during think-aloud sessions on diagnostic essays and in interviews, and as they reported final grades and assessments, the teachers offered their impressions of the relationship of placement decisions to course work.

The Findings and Recommendations of This Study

The study found that placement readers at Cardinal College see students as existing on a continuum of ability linking reading and writing, as the program asserts, and they indeed use teaching specific criteria to determine a student's position on that continuum, presenting a pedagogically useful view of students' placement essays. The placement criteria, which focus primarily on how students come to and represent a reading of complex texts through writing, expose students who, past experience and the current research showed, may go on to struggle in a mainstream college writing course because of particular combinations of reading and writing behavior. However, the study also found that because of the change to Cardinal College's curriculum with the loss of Course 11A, the expert-reader model (Smith 1992/2009, 1993) that placement readers previously operated from was no longer being followed as the Course 11 reader split her focus between her recent Course 11 teaching experience and her long-standing knowledge and expertise on students in the eliminated Course 11A.

As the Course 11 reader recognized behaviors of students in the eliminated course, she stopped making a decision only about Course 11, and rather than push the student down to the other reader, she instead weighed the merits of both possible existing course placement options for herself, often making a notation on a student's file suggesting the advising office register the student in the section of a particular teacher who had experience teaching the old course. In short, the expertise of the Course 10 reader was not being consistently engaged to help discern which students would be most or least concerning. Yet, the anxiety of the Course 11 reader was not unfounded. Pushing students down to the six-credit course option, when it may not be necessary for all

students, could have consequences for students' progress through their courses of study and their engagement with the learning in the writing program, not to mention cause pushback from students, other departments, and the administration. Further, an analysis of the course criteria and outcomes showed that Course 11 teachers with different past teaching experiences—either Course 11 or old Course 11A—have different understandings of students' reading difficulties that could potentially affect how they respond to students' reading issues. On either side, the Course 11 placement reader faces the possibility of a less than adequate placement decision. But the current practice of recommending troubling students to only some teachers will exacerbate the differences between the two subsets of Course 11 teachers and is not a long-term solution.

It is the recommendation of this study first that Cardinal College's writing program must more formally and structurally address the loss of the old Course 11A, reimagining the work of the remaining courses to address the needs of the Course 11A portion of the student population. And second, in the interim, the work of placement and teaching must be modified. For placement, Smith's (1992/2009, 1993) expert-reader model must be reasserted, with a modification made by Haswell and Wyche-Smith (1994) that has readers flag students who do not fit neatly into their existing course and wait to make decisions on these students in concert with other readers. This process will engage the expertise of the Course 10 reader and produce conversations about troubling student work that the writing program can use to redevelop the existing courses, meeting the needs of these students. In regards to teaching, course teachers must be encouraged to seek out conversation and collaboration as they teach students during the transition to Cardinal College's courses. By turning to each other, teachers can more fully recognize

and strategize for students in subpopulations with which they are less familiar, and they can generate and test out approaches and solutions to specific students and scenarios.

The Value of Validation Research

Having outlined the focus, research questions, data, findings, and recommendations of the present study, I want to offer some statement of the value of this work. As I see it, there are two levels of value to the validation research undertaken in this study. First, the study had value to the actual site since it helped the writing program at Cardinal College address its concerns and it contributed to the work of developing the program. Second, such work can mean a great deal for the field of composition and the work of writing assessment, in particular, through what it offers those with shared contexts and concerns, what it shows about curricular change, and how it cultivates a culture of teaching and learning based writing assessment.

The Value for Cardinal College

Validation work provided the program with meaningful understandings of its courses, and an in-depth, qualitative view of its teachers and students. Additionally, validation research has helped the program start to investigate its placement assumptions. Through this study, Cardinal College has been able to consider the way it meets the needs of its particular students, including those students who used to fit into the discontinued course, and how it might use future placement reading and departmental work to adjust the courses to better meet students' needs. Thus, the current study has had a direct and potentially profound effect on the work of Cardinal College's writing program even as the study has revealed the need to reevaluate the courses and improve the placement reading process. Having the clear and detailed description of the placement and course

determinations and the many student examples presented here will serve the program well as it decides how best to reconfigure the courses, and as it reestablishes placement reading around new courses boundaries.

The Value for the Field

The description and discussion of Cardinal College that was produced through this study allowed the program to study itself—to check the effectiveness of its procedures and the validity of its decisions—by looking closely at how the criteria under discussion by the teachers in the program related but did not fully align with those under discussion during placement. Yet additionally, this work, as I see it, was valuable in three other ways, each larger than the direct benefit to Cardinal College specifically.

Those with similar contexts and concerns. Since validation research requires that the interpretive argument that is sketched be specific to the particular site and context, this study describes in detail the College, its courses, its students, and its teachers, as well as the goals and vision of its writing program. For those who share similar teaching and learning contexts or concerns, such a detailed description and discussion can be instructive or helpful to their own situations. More specifically, this study would be helpful in three other teaching and researching contexts. First, for other small liberal arts colleges who primarily teach writing to undergraduates and who wish to be able to conduct scholarship—particularly assessment scholarship—that doesn't take away from their primary work of teaching but instead feeds it (Donahue & Moon, 2007; Gladstein & Regaignon, 2012), the study here provides a procedure and model for studying your own practices and improving teaching and learning goals through ongoing validation.

Second, for those writing programs with integrated reading and writing courses, a great deal can be learned from the criteria developed in Chapter Four of this study—particularly the descriptions provided of the way teachers understand students’ misreading, how they theorize students must conceive of the purpose or tasks of reading, how they weigh the importance of reading issues against writing ones, and how they imagine the impact reading and writing have on each other. Relatedly, those studying how students’ reading and writing development impact each other can work with and add to depictions of developmental continuums on integrated reading and writing criteria as well.

Third, for writing programs who are worried about some type of disagreement or misalignment in the context of their site, this study provides a glimpse at how to use validation to study existing misalignments. The example of Cardinal College’s misalignment due to shifting course boundaries brought about by unplanned curricular change, and the way that misalignment was uncovered can help others make sense of their specific institutional histories or teaching differences which may contributed to their own misalignment or disagreement in some form.

Curricular change. The findings uncovered suggest something about the importance of the work of assessment during times of curricular change. The study of Cardinal College’s specific situation, with its example of a program making placement decisions after the elimination of a course offering, is particularly instructive on the importance of validation work in times of curricular change generally. As assessments are only valid in the context of their intended use, and as we are responsible for checking that the assumptions we’ve made in the design and implementation of our assessments

are aligned with reality, the process of curricular change requires substantial validation work. Yet the usefulness of validation in times of curricular change, particularly externally required change, is deeply contextual as well.

It is difficult for teachers and programs to know, when faced with such changes, what they must do to best meet the needs of their students. It is possible to have teaching knowledge and expertise that will find no outlet when the courses themselves are not aligned with what teachers have come to know about their particular student population. Forced curricular change in particular can make it hard for teachers to engage with their expertise. In these moments, teacher's existing expertise and the key understandings contained in their conceptual frameworks, built over many years, can seem more like a burden when they only help teacher's note pedagogically relevant issues that can find no clear curricular outlet.

It is in these moments that teachers and programs must fight to make decisions that improve teaching and learning, even when their choices are not ideal. Teachers must think about what work can best help them meet the needs of their students in the face of existing contextual constraints. They must consider which decisions will allow them to preserve their ability to do work they believe in. Validation work can help teachers develop contextualized understandings and interpretations, and continue engaging with their expertise in order to figure out how to find an outlet and expression for what they know. Through validation, teachers can identify what isn't working, gather support for their expert interpretations, and make changes.

Culture of teaching-and-learning-based writing assessment. The study of specific sites like this one, with published results discussing the work toward validating

the decisions being made, helps in the creation and promotion of a culture in composition where writing assessment is done with a focus on locally controlled, site-based assessments and their particular teaching and learning goals and contexts. Current writing assessment theory (CCCC, 2009; Huot, 1996/2009, 2002; Huot & O'Neill, 2009) implores us to make a case for, through our own examples, the kind of writing assessment that aids the real work of teaching and learning writing. Yet, the publication of such studies is comparatively rare in the literature. This study adds another example of such work to the literature of writing assessment. As it presents a published example of a locally-created writing assessment and the work of validating it, this study reasserts that assessment, to be well done, must be deeply connected to the specific teaching goals and contexts it is assessing. Examples like this one can challenge the mythoi of standardized assessments, decontextualized writing, and one-size-fits-all teaching and learning, and they can make it possible for other teachers and programs to study and maintain the integrity of their work despite external pressure.

An Overview of Upcoming Chapters

As a final move, I want to offer a breakdown of the structure of upcoming chapters so their purpose is clear. In Chapter Two, I offer a review of relevant writing assessment literature to present a history of writing assessment in which important shifts and key connections are revealed. Specifically, I first offer a more detailed description of the conflict between those who study testing and those who teach writing and how that conflict evidences real differences of concern, focus, and value. From there, I discuss some of the literature in educational measurement on the changing definition of validity across time. My goal with these two moves is to draw several parallels between the

disciplinary and teaching concerns of English teachers and the changing definition and theory of validity.

Next, I move forward through writing assessment history—in particular the history of holistic scoring—to argue that the parallels between teachers’ concerns and new expressions of validity could only be realized when holistic scoring and the focus of reliability was questioned sufficiently to initiate a new central concern with validity in writing assessment. Given this new concern with validity in the literature, I go on to detail Smith’s (1992/2009, 1993) studies to establish their link to the principles of newer definitions of validity and to discuss Smith’s discoveries on the importance of teacher expertise and local context. I argue that his discoveries reveal the importance of writing teachers’ involvement in validation. Finally, I move to more recent literature on placement testing specifically to discuss the significance of a teaching and learning based assessment built on validity, teacher expertise, and the local context.

In Chapter Three, I detail the specifics of the design of this study. I begin with an explanation of what it means to validate according to current theories since I believe that in order to make convincing arguments to various stakeholders, we must know the language and practice at work in testing theory so that we can use or adapt it for our work. Next, I provide details of the particular context of Cardinal College as I sketch the interpretive argument for its writing placement process, including a discussion of the College population, courses, teachers, placement prompt, and placement reading process. Focusing on the key assumption of the interpretive argument under investigation in this study, I discuss why this assumption was the most pressing for study and how future studies must continue the work of validating Cardinal College’s program by examining

other pieces of the interpretive argument. Once the interpretive argument and key assumption are sketched, I use the sketch to provide a rationale for the evidence required in this study and for the data collection and analysis decisions made in the design of the study. After, I describe chronologically the process for data collection, analysis, and coding. Finally, I discuss the strengths and limitations of the data collection and analysis.

In Chapter Four, I present the findings of the study. The analysis of the placement criteria and the concerns of readers revealed first that Cardinal College's placement readers see students on a continuum of linked reading and writing ability, just as the program and courses posit. Second, the exact border between the two courses, Course 10 and Course 11, was impacted by a reader's sense of the potential for a student to experience reading difficulty. Third, that potential was hard for readers to determine when the single piece of student's placement writing had characteristics that provided insufficient evidence for the reader to feel confident about their determination on the strengths and limits of a student's reading ability.

Thus, the readers frequently used teaching specific knowledge expressing pedagogical concerns centered on how to teach students with particular reading and writing behaviors to determine where problematic students might struggle or excel, as they should given Smith's (1992/2009, 1993) model. Yet, this knowledge often engaged readers' previous experiences teaching students in a now discontinued course that fell between the two existing courses, Course 11A. As a result, Smith's expert-reader model was not being properly followed as the Course 11 reader worried about the potential for misplacement given the loss of the old course, and her attention was divided between considering both course options for old Course 11A student essays rather than making a

decision for one course. To address her worry, the reader often placed old course students with teachers who had experience teaching the old course, and this decision has the potential to exacerbate existing differences in teachers' understandings of the student population.

Furthermore, the analysis of the course criteria and the concerns of course teachers revealed that Course 11 teachers were divided into two groups based on their beliefs about the necessity of assessing the incoming reading abilities of Course 11 students, and that divide coincided with previous Course 11 teaching experience. Those who once taught Course 11A attempted to discern how much and in what ways a student would struggle to read the kind of texts in the course, and those who hadn't primarily taught Course 11A believed incoming students would struggle in similar ways addressed already by the course. Therefore, placement designations were not aligned with all Course 11 teachers. However, Course 10 teachers, in contrast, had a shared set of criteria and vision for their course, and a strong sense of the specific reading habits of underprepared students valuable for distinguishing less severe reading habits and behaviors from more severe ones, which aligns with placement decisions made regarding Course 10.

Finally, an analysis of all data, including teachers ranking of students' fit to the course, as well as final grades and course outcomes confirmed the earlier finding that Course 10 teachers' designations and decisions aligned with placement ones while Course 11 teachers' designations and decisions were not fully aligned. However, the course outcomes showed that there must be some relationship between Course 11 placement decisions and Course 11 since students flagged at placement as old Course

11A students underperformed in relation to their Course 11 peers in terms of course grades. Finally, since Course 11A students also outperformed their Course 10 peers, the program must consider the negative consequences old Course 11A students could experience in either placement, as students face demotivation on one end and the potential for insurmountable difficulty on the other.

In Chapter Five, I go on to consider what the findings of this study mean for Cardinal College and the field. In regards to Cardinal College, I recommend that since the return of Course 11A is unlikely, the program has to focus its future departmental meetings and staff development on reimagining the program without Course 11A, sharing experiences to develop a consistent vision, and developing strategies for use with weak readers in the program, all while they teach through the transition to the courses. Further, I recommend that the placement reading process at Cardinal College should be reestablished around Smith's expert-reader model, but revised to encourage readers to admit reservations and save concerning student essays so that decisions can be made in conversation with the other course reader (Haswell & Wyche-Smith, 1994). Finally, I discuss how such a process can uncover key issues that will aid departmental decisions about the boundaries of these courses, and which will result in necessary staff development and the reimagining of teacher's course expertise.

In regards to the field, I first discuss how this study could aid those who share similar teaching and learning contexts in thinking about the work of their own validation—specifically small liberal arts colleges who teach writing to undergraduates and desire scholarly research that impacts their primary work of teaching, programs with integrated reading and writing courses or those studying the interconnectedness of

reading and writing development, and those who wish to study the misalignments and disagreements in the context of their programs. Next, I detail the lessons learned here regarding what it means to teach and assess writing, in this case make placement decisions during times of curricular change, particularly external curricular change. Most specifically, I discuss how teachers feel when they can find no outlet for what they know about their student populations, and I suggest that validation presents a way for teachers to find their way through, focusing on their students and the integrity of their work, while they fight to continue making teaching and learning based decisions useful to their contexts. And finally, I talk about the value of this study as an additional example in the literature of locally controlled, site-based writing assessment focused on teaching and learning (CCCC, 2009; Huot, 1996/2009, 2002; Huot & O'Neill, 2009), and how the addition of examples like this one can create a culture in composition of teaching and learning based assessment that make it possible to challenge the mythoi of standardized assessments, decontextualized writing, and one-size-fits-all teaching and learning.

Conclusion

Writing assessment theory suggests that assessments should be locally-controlled, site-based, and contextualized (Huot, 1996/2009, 2002). Yet, outside of those pieces of the literature of writing assessment that have been specifically written by writing teachers, such concerns are not given primary accent. If the work of teaching and learning is going to be the emphasis of assessment, the theory of writing assessment that is common within composition must be communicated to other interested stakeholders. To do so, the long standing divide between writing testing and writing teaching must be breached. Huot (2002) has asserted that validation theory can be employed by writing

teachers to bridge the divide in writing assessment. This study was conducted to begin the process of using validation theory to focus Cardinal College's writing assessment on its teaching and learning concerns and goals; yet it also presents the results of its study to reiterate the importance and centrality of teaching and learning concerns in the work of writing assessment generally.

CHAPTER TWO

REVIEW OF LITERATURE

When a student is accepted to a college or university, one of the first concerns for all involved is what classes the student will take. Most institutions today have a few different course options, and these include one or more remedial or basic writing options (Elliot, Deess, Rudiny, & Kamal, 2012). Some scholars have claimed remedial courses are offered because high schools are not preparing students adequately for college writing work (Gere, Aull, Green, & Porter, 2010; Mattern & Packman, 2009). Elliot, Deess, Rudiny, and Kamal (2012) have suggested that this may be a perception that is potentially untrue since on measures like the National Assessment of Educational Progress (NAEP), students' scores have increased over time.

While there have been critiques of remediation and basic writing courses (see for instance, Adams, 1993; Crowley, 1996), and critiques which have suggested that testing and remediation are businesses which capitalize on the creation of tests and course materials for remedial courses (Elliot et al., 2012), a majority of teachers who work with underprepared students in basic writing courses would agree that these students do face difficulties. Historically, basic writing teachers can name exactly what makes a basic writer a basic writer (Bartholomae, 1979, 1985/2005, 1987; Greenberg, 1987; Shaughnessy, 1977), at least within the context of their students and programs (see Troyka, 1987 on context and basic writing). These teachers would agree that, despite concerns in the literature about what Otte and Mlynarczyk (2010) have described as “ghettoization” (p. 67) (see for example Bartholomae, 1993; Crowley, 1996; Shor, 1997),

there are real benefits to additional course work and preparation for basic writing students before going on to college composition.

While a majority of basic writing teachers would agree that there are benefits in basic writing instruction for those students who need it, teachers who are writing assessment specialists have argued additionally that adequately placing students into basic writing should be a chief concern for those directly knowledgeable and involved in the teaching of these students (Greenberg 1993; White, 1995c). Placement is clearly an ethical issue given the real social consequences that are a part of it (Messick, 1981, 1989a); remedial courses often do not count for college credit, they take additional time to complete, they may not count toward financial aid totals, and they can often carry a stigma. This has created a real need for writing program directors and teachers to check the *adequacy* (Smith, 1993, p. 143) of their decisions about which students would benefit from remedial or basic writing courses and which would not.¹

As outlined in Chapter One, the focus of this research overall was on studying the validity of the placement process at a small, private college in the suburbs of a mid-sized, Mid-Atlantic city which uses Smith's (1992/2009, 1993) *expert reader model* to place students in reading and writing integrated courses. This study answered the following research question:

¹ It is important to note at this point that I am purposely not using the word "accountable" as I discuss concerns about placement. On the surface, accountability is the perfect word since it signifies a responsibility inherent in a position of expertise and authority. However, while accountability may be used with the intention of improving education, the word as it is used currently in reference to education suggests a meaning now largely characterized by a portrayal of teachers as incapable of responsibility without heavy top-down monitoring. Thus it works to strip teachers of their expertise and authority—two things essential for anyone to be truly "accountable"—as it paints teachers as disinterested in getting things right, in checking their decisions, and in being transparent and responsible. I wish to avoid suggesting what I emphatically believe to be untrue.

- Do expert readers at Cardinal College use curriculum/course related and teaching specific criteria to offer a pedagogically useful and purposeful, view of students' placement essays? Are expert readers' criteria (and by extension their decisions) linked to course criteria?

To answer this question, it was necessary to break the research question into four subquestions:

1. What are expert-readers' placement criteria?
2. Are expert-readers' placement criteria teaching specific?
3. Are expert-readers' placement criteria linked to course criteria?
4. Are expert-readers' placement concerns and decisions linked to course concerns and decisions?

Pursuing its research question and subquestions, this study began the process of crafting a validity argument for Cardinal College's placement by specifying the program's *interpretive argument* (Kane, 2006, p. 22) and investigating a key assumption of the writing program's argument—that placement readers place students according to course content criteria.

As this study began the validation research on the placement process at Cardinal College, it rested on a definition of validity that takes as a starting point the need for ongoing study into an assessment process, with the goal of crafting an argument for the adequacy and appropriateness of an assessment, mindful of decisions to be made using the assessment and the consequences of its use (Cronbach 1988; Kane, 2006; Messick, 1989a, 1989b). While Chapter Three deals more directly with how to validate, the particulars of validation for Cardinal College, and the line of inquiry in this study, this

chapter reviews relevant assessment literature to craft an argument for why we—writing teachers and WPAs—must be responsible for the validation of our own programs and how that validation effort can be served by an understanding of the history of conflict in writing assessment between those working in educational measurement and ourselves.

In this chapter, I first recognize the reality of the conflict between those who study testing and those who teach writing, and how that conflict evidences real differences of concern, focus, and value. Second, I turn to the literature in education measurement to review the changing definition of validity across time. Third, once I have outlined these changes, I draw parallels between the disciplinary and teaching concerns of English teachers and these changing definitions to make clear that as the definition of validity (and its relationship to reliability) in educational measurement has changed, the term of validity has developed in ways that have increasingly come closer to addressing the historical and current concerns of English teachers and WPAs. Fourth, I argue the parallels between teachers' concerns and new expressions of validity could only be realized when holistic scoring and the focus of reliability was questioned sufficiently to initiate a new central concern with validity in writing assessment, and to raise questions about rater agreement. I argue this new central concern with validity brought about an acknowledgement of the contributions of those in educational measurement who had refined their understanding of validity, and through that, an acknowledgement of the contributions of English teachers who were not solely focused on reliability in their assessment work. These realizations, I argue, have brought about what Huot (2002) has called a “(re)articulating” of writing assessment for teaching and learning (p. 3).

Fifth, I focus on the moment in time when factors were ripe for validity to move to the forefront of assessment work (instead of reliability) by looking at the work of Smith (1992/2009, 1993). I show that centering on validity made it possible for a realization of the importance of Smith's work. Specifically, I detail Smith's studies, establish their link to principles of newer definitions of validity, and discuss Smith's discoveries on the importance of teacher expertise and local context. Sixth, using Smith's focus on studying his program and his findings on teacher expertise and local context, I examine specific literature on rater agreement and writing prompts to argue that validation by writing teachers is especially important for teaching our students and developing our field due to our special expertise in teaching and learning. Finally, I move the significance of a teaching and learning based assessment built on validity, teacher expertise, and the local context to the current moment in time and the concern of placement specifically. By discussing specific examples of literature that seem to ignore these lessons, I hope to defend my reading of writing assessment history as having current relevance.

Two Sides of Writing Assessment

In writing assessment, there have long been two groups working—those in educational measurement (such as those people associated with ACT, ETS, the College Board, or the field of psychometrics) and English teachers (and later, as the field developed, composition teachers). Prior to the 1960s, the work of assessment was done primarily by those studying testing specifically (like figures from ETS), and these people had little experience in the discipline or teaching of English (Elliot, 2005). As part of test development, they perhaps would hire or consult with English teachers, when it was felt

that discipline specific knowledge was necessary, but it was not until much later, with Paul Diederich, where someone working for ETS had significant English experience and accented that knowledge (Elliot, 2005).

At times, the situation of writing assessment has been characterized as a war. For example, White (1993) described those on either side as enemies, and his narrative painted a picture of an epic battle. War is a particularly telling way to characterize the division since these two different communities historically have had conflicting concerns, different focuses, and solutions that emerge from their differing values or conceptions of writing (Williamson, 1993), and both sides at times have either ignored the contributions and theories of the other or have struggled to discuss emerging issues in ways that would solve the concerns of all.

Many of the differences between educational measurement specialists and English teachers are outlined in the influential *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations* edited by Williamson and Huot (1993). Specifically, the piece by White (1993) describes differences in both theory and ideology between these groups. He claims that “reductive concepts of reading and writing” were at work in indirect measures, equating correctness with writing ability (White, 1993, p. 81). Further, he acknowledges in his recounting of his support for holistic scoring that the human aspect of community-building involved in the direct reading of students’ work was ideologically appealing to teachers but unimportant to testers. Camp (1993) outlines differences in the views on reading and writing expressed by those in educational measurement and those who teach writing (with those in educational measurement seeing writing as an expression and not a construction of

meaning). Williamson (1993) traces the history of psychometrics steeped in behaviorist theories that “did not keep pace with developments in cognitive psychology” (p. 6), and thus did not deal directly with unobservable conceptions of the mind. He also outlines the psychometric desire for abstract truth and objectivity over an understanding of the role of context in writing.

In response to a building acknowledgement of the differences between those who teach writing and those who traditionally assessed it, Huot (1996, 1996/2009, 2002) tried to look across the divide between these groups, asking what we could learn from each other. He discussed a new definition of validity in the educational measurement field that he felt would no longer conflict with the English teacher’s vision, and instead would allow English teachers to make a case for the kind of writing assessment they had been asking for all along. Instead of framing the division as a war, with educational measurement specialists as enemies, as White (1993) had done, Huot (1996, 1996/2009, 2002) focused on the reality that there were different stakeholders who each had some concern invested in the work of writing assessment, and he asked writing teachers to think about the importance of talking across these divisions.

Changes in the Definition of Validity (and Its Relationship to Reliability)

To truly see the importance of the newer definition of validity that Huot (1996) found as he studied educational measurement literature, a review of the history of validity is necessary. For his chapter in the influential *Educational Measurement*, Kane (2006) described different historical conceptions of validity and outlined key texts useful both for seeing the different definitions of validity and for recognizing those who were responsible for these definitions. In particular, he highlighted the work of Cronbach and

Meehl (1955), Cronbach (1971, 1980, 1988, 1989) and Messick (1975, 1981, 1988, 1989b), as well as his own (Kane, 1982, 1992).

Using Kane's (2006) description and a reading of these additional works, I have outlined seven stages in the development of validity that are useful to understand in relation to the concerns of English teachers.² First, validity was initially expressed as criterion validity (in two forms: predictive and concurrent). Second, content validity was added to criterion validity. Third, construct validity was introduced specifically for dealing with theories. Fourth, validity solidified as having three forms (or four if you split criterion into predictive and concurrent) that could be selected for specific purposes. Fifth, despite the persistence of the three (or four) types of validity, there was a growing desire for a unified theory. Sixth, validity was unified under the concept of construct validity. Seventh, validity as an argument emerged. Once finished outlining these seven stages, I go on to discuss briefly the relationship of reliability to validity since an understanding of the relationship between these two terms impacts the history of writing assessment.

Criterion Validity

Between 1920 and 1950, criterion validity was the preferred conceptualization of validity (Kane, 2006). Criterion validity was basically an assessment of how closely a test aligned with performance on an actual task by either seeing how the measure predicted performance on that task or how it aligned with other measures of that task. Thus, there were two kinds of criterion validity—predictive and concurrent. In the case of placement

² While I emphasize the progression of changes and development to the definition of validity as I outline these seven stages, it is important to mention that recent validity research studies can be found that use older theories of validity as if the understanding of validity never progressed past certain moments in time. In spite of the presence of recent studies using outdated theories, I speak of older theories as existing in the past and characterize these stages as having defined boundaries based on the literature of validity theory.

testing, predictive criterion validity could align the score on the placement test with future course grades while concurrent criterion validity might seek to align the score on the placement test with other concurrent measures like SAT scores or high school GPAs.

Kane (2006) claimed that the problem with criterion validity was not in aligning the measure with the chosen criterion (which he suggests is easy); the problem was in choosing the criterion. All criteria would be “questionable” (Kane, 2006, p. 19) since the criteria themselves would need to be validated. As Kane discussed the limitations of criterion validity, he made it clear that choosing the criterion (and validating that choice) was not part of the model of establishing criterion validity, and that was the problem. The criterion was accepted as an adequate measure without any justification of that subjective judgment.

Adding Content Validity

Kane (2006) showed that the content model was added to the criterion model to deal with deficiencies in using the criterion model alone. Content validity focused on how the test represented some set of the skills involved in the actual task and on getting an adequate sample taken fairly. Thus, it seemingly helped solve the problem of criterion validity because without content validity, one could infinitely align one measure to another with no way to establish that the criteria were reasonable.

While content validity helped solve problems of establishing criteria, Kane (2006) revealed that the problem with content validity was also its subjective nature: “Content based analyses tend to rely on judgments about the relevance and representativeness of test tasks. When these judgments are produced by test developers, they have a natural tendency to confirm the proposed interpretation” (p. 19). In short, those developing tests

would assert that their test tasks were representative of the actual task (or would find experts who would) rather than pursuing other possible explanations. In addition, Messick (1989b) claimed that content validity also would not work for theoretical ideas about cognitive processes since these involved only theories of what was going on in the mind, which was not confirmable through the established procedure of content validity.

Introducing Construct Validity for Theories

In the early 1950s, the American Psychological Association (APA) was seeking to offer guidance on the validation of theoretical concepts (Kane, 2006). The development of construct validity to fill this need was summed up by Cronbach (1971) as he recalled the reasoning behind its creation. Using the example of developing a personality test of ego strength, Cronbach (1971) explained that since ideas like ego strength had “no...criterion to predict” and also lacked “a domain of content to sample,” all that remained was a theory of what ego strength was (p. 462). Therefore, to determine validity, one had to test out how scores on measures would “conform to the theoretical expectations” (Cronbach, 1971, p. 463). Cronbach further explained the construct validation process in terms of nomological networks where constructs—theoretical ideas about what could not be observed directly—were tied to observable data that should play out in ways predicted by the theory of the construct. Cronbach upheld that networks developed as sets of constructs and observable data were linked together such that one could become more and more confident in a measure when one piece of data could help one infer the value of others in the network.

As Cronbach (1971) made clear, construct validity dealt with finding evidence to substantiate that a theory predicted observable behaviors or performances. Drawing from

Suppe (1977) on the hypothetico-deductive model of scientific theories, Kane (2006) explained further:

[I]f the predictions derived from the theory do not agree with observations, then either the theory is wrong, the measurements are not appropriate indicators of the constructs in the theory, or some ancillary assumption was violated. If diverse predictions based on the theory are confirmed, the theory and the interpretation of scores in terms of the theory are supported. (p. 20)

Kane's explanation, taken with the nomological network described by Cronbach (1971), not only reveals the focus of construct validity on outlining theories clearly, but also shows how construct validity was a system of theory testing and building based not only on the creation of theories but on how observable information corroborated proposed theories.

Solidification of Four Kinds of Validity

What emerged from the addition of each new kind of validity—from criterion to content then to construct— was an understanding of validity divided into three different types. Determining which kind of validity to use was a matter of understanding if your particular case involved theories or something observable. The general practice was to use construct validity in relation only to the validation of theories. Kane (2006) made this clear as he characterized the 1966 American Psychological Association, American Educational Research Association, and National Council on Measurement in Education *Standards for Educational and Psychological Tests and Manuals*, and even the 1974 ones, as showing a view of validity which held that there were four kinds of validity in total (concurrent, predictive, content, and construct) useful for different situations of

validation—observable or theoretical. Interestingly though, pieces written prior to these standards began to suggest another possible approach with construct at the center of validity, but Kane (2006) argued that this idea was not represented in the standards and was not common place in usage.

Desire for a Unified Theory

According to Kane (2006), Loevinger (1957) appears to have first suggested that construct validity was really the whole of validity and that criterion and content validity were for secondary special purposes. Even though Cronbach and Meehl (1955) first expressed construct validity, they did not intend Loevinger's meaning (Shepard, 1993; Kane, 2006). To Cronbach and Meehl, criterion and content validity would work for the validation of observable performances of tasks, but construct validity was used in dealing with a theoretical conception proposing what happened in the mind which was unobservable in nature. However, Kane considers Cronbach and Meehl's actual words to make clear that even though they did not put forth construct validity as a central or unifying concept (this would begin to emerge later in Cronbach, 1971), Cronbach and Meehl did suggest that working on the construct was valuable to all tests because doing so, in the words of Kane (2006), would give us a "deeper understanding of the construct or processes accounting for test performances" (p. 20).

Cronbach's (1971) piece suggested, although he did not directly call for, a unified view of validity in part because it asked for multiple kinds of evidence and not a picking and choosing of evidence (as was the case with the different types of validity already present). While the late 1970s to early 1980s were colored with the push for a unified theory (Cronbach, 1980; Guion, 1980; Messick, 1975; Tenopir, 1977), this time period

also showed a desire for clear guidance on what kinds of evidence would be needed to validate (Kane, 2006). Therefore, while there was an expressed need for some kind of overall guide, the desire for clear forms of evidence seemed to keep the old terms and definitions of validity alive. Still, there was a growing realization that some unified guidelines would have to emerge.

A Unified Theory Under Construct Validity

The official presentation of a unified theory came about in the 1985 American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) standards (Kane, 2006). The unified theory took construct validity as its central concept. Construct validity as the central matter of validity is perhaps best seen in Messick (1988, 1989b) where construct was put forth as the primary concern while content and criterion models operated only in supporting functions secondary to the work of specifying interpretations, assumptions, and alternative views for all measures.

Kane (2006) claimed there were three positives to taking up construct as the unified approach. First, the model opened up validity to consider “a broad array of issues inherent in the interpretations and uses of test scores” which supplied a welcome complexity to criterion validity (Kane, 2006, p. 21). One such issue was the consequences of tests which Messick (1989b) placed in equal weight to the construct. A second improvement, according to Kane (2006), was the focus of unified validity being rightfully placed on “assumptions in score interpretations” that would need to be “checked” (p. 21). Instead of ignoring these assumptions, they would have to be spelled out and validated as well. And third, Kane (2006) accented how “the construct model

allowed for the possibility of alternative interpretations and uses of test scores” (p. 21), which was positive because it reduced the possibility of test developers confirming already held assertions with no consideration to other interpretations.

Validity as Argument

Because of construct validity’s origins for working exclusively with theories, Kane (2006) asserted that it was important to move past its existing guidelines in order to provide sufficient guidance for all validation. As Kane discussed Cronbach’s (1989) description of validation as being infinitely ongoing, he detailed the concern that construct validity within its old explication did not give us much guidance on what was *enough* validation. What is adequate? If we have no clear guidelines on what is enough, then we may rely again on our own inclinations. For as Cronbach (1989) asserted, while we are inclined to point out the flaws in others’ interpretations, we are reluctant to do the same with our own.

Kane (2006) used the literature on construct validity (specifically, Cronbach, 1971, 1988; Cronbach & Meehl, 1955; Messick, 1989b), and he outlined three principles gleaned from the expression of construct validity that created guidelines for all validation free from their origins in relation to the validation of theories: “the need for an explicit statement of the proposed interpretation, the need for extended analysis in validation and the need to consider alternative interpretations” (p. 22).

These principles frame validity, as understood in current theory, as the crafting of an argument, and they are the expression of validity characterized by the 1999 AERA, APA, & NCME standards. The argument approach, first put forward by Cronbach (1988), is perhaps best explained by Kane (1992):

The argument-based approach to validation adopts the interpretive argument as the framework for collecting and presenting validity evidence and seeks to provide convincing evidence for its inferences and assumptions, especially its most questionable assumptions. One (a) decides on the statements and decisions to be based on the test scores, (b) specifies the inferences and assumptions leading from the test scores to these statements and decisions, (c) identifies potential competing interpretations, and (d) seeks evidence supporting the inferences and assumptions in the proposed interpretive argument and refuting potential counterarguments. (p. 527)

The argument frame provided here details that all assumptions, interpretations, and decisions involved in a test's use must be made clear, investigated for alternative possibilities, and backed up with evidence.

Although Kane (1992) extrapolated principles from earlier expressions of construct validity to frame validity as an argument, Kane (2006) clarified that there are a few key ways that argument, as a frame for validity, accents or alters past beliefs, practices, or issues. First, a focus on argument accented the claim of Cronbach (1971) that validity is not in the test itself but in the use of the test in a given situation; seeing validity as an argument emphasized the judgment and reasoning behind the use of the test in a particular situation. Second, the guidelines for crafting a validity argument, while uniform, will no longer result in identical validation research or a uniform process of validation since the evidence that is required will vary based on the “specific inferences and assumptions” related to a particular “application” (Kane, 2006, p. 22). Third, rather than having a set or acceptable stopping point, validation work is infinitely ongoing, as

we can never fully prove an argument. However, Kane suggested that instead of being confused as to what constitutes enough validation, as we were with past definitions, the value of using argument as a framework for validity is in how it helps focus us on what would be the biggest assumptions of, problems with, or challenges to our arguments. While we may never be *finished*, we can determine what evidence and explanation will be sufficient or necessary to convince others of the reasonableness of our claims.

Today, when discussing validity, many cite the definition offered by Messick (1989b), that validity is “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment” (p. 13, emphasis in original). His definition offers a fair summation of the facets of validity that have been discussed, and brings them into current understanding. In his definition, validity (and knowing when we have done enough to argue for it) depends on many things: the evidence, the theoretical support, the assumptions we make, the specifics of our particular situation, the particular decisions we will make using this test, etc. Further, Kane’s (2006) expression of validity as an argument lines up well with the foregrounding of validity as a “judgment” of “degree” expressed in Messick’s definition. Researchers may have done enough work to make some claim for validity, but to increase that degree, our work often will continue into other facets we have not yet explored.

The Relationship of Reliability to Validity

In classical test theory, there is a close relationship between validity and reliability (Cherry & Meyer, 1993), and these terms have been discussed together at times in the literature of writing assessment from those working in educational measurement

and from English teachers (see for example, Camp, 1993; Greenberg, 1992; Huot 1990a; Moss 1994; Wiggins 1994). Since both terms are discussed going forward, it is necessary to give a basic explanation of reliability to compliment the discussion of validity.

Although, since the focus in this study and in the argument I am crafting in this review is about validity primarily, the discussion of reliability will not be as detailed as that for validity.

Reliability has generally been expressed as the consistency of a measure, and this consistency has a lot to do with the concept of a true score—the idea that there is a score that represents a person’s ability perfectly (Huot, 1990a). Reliability has therefore been related to an assurance that very little error in measurement has arisen based on test attributes, test conditions, rating procedures, or student characteristics that are not related to those being measured (Cherry & Meyer, 1993). Typically, this has meant that a score on one measure can be replicated in retests (or on equivalent questions) or by different raters (interrater reliability) (Cherry & Meyer, 1993).

Despite reliability’s expression as a separate property of psychometric test theory, it has also been discussed through its relationship to validity. Throughout most of the changing history of validity, reliability has consistently been expressed as a “necessary but insufficient condition for validity” (Cherry & Meyer, 1993, p. 110). That is, to claim validity, one would need to determine reliability as well. And even in the argument framework, this relationship between validity and reliability is solidified since an argument that asserts the valid use of a test would require some assurances that its use is consistent and fair. Yet, among English teachers in particular, as I show shortly, there has been great concern about the emphasis placed on reliability.

In addition to issues with the relationship of reliability to validity, there have been discussions of the problems with how reliability is calculated (Cherry & Meyer, 1993; Parkes, 2007). When examined, these discussions reveal the limits of reliability for the work of writing teachers. Take for instance Parkes (2007), who attributed problems in calculating reliability to a confusion of reliability's values—“dependability, stability, accuracy, and consistency” (p. 2)—with the methodologies employed. In calling for reliability's expression as an argument pertaining to test use, just as validity has come to be expressed, Parkes began an interesting discussion of what the concept of reliability does not acknowledge. One important example was detailed as he showed that the methodology of establishing reliability asks for specific “sources of evidence” that do not necessarily line up with certain situations (Parkes, 2007, p. 2). For instance, he discussed the example of a classroom teacher having her students use a rubric on collaboration to assess each other, and he showed how in classical theory, since the students are learning about collaboration as they use the rubric, they are unreliable raters. He discussed how using an argument framework will allow the field to develop new evidence and explanations that would be acceptable in a scenario like this one—where the teacher could make a case for reliability based on how she gives more weight to later ratings and how she asks students to back up their ratings with details. So while many people working in writing assessment would argue that validity is related to reliability, here there is an acknowledgement of the limits of reliability for validation work.

Parallels Between Validity and English Teacher Concerns

Tracking validity's redefinition across time, we see that it has increasingly addressed the concerns of English teachers. Going forward, I discuss three specific

concerns expressed by English teachers—sacrificing validity to reliability through the indirect assessment of writing; constructing assessments that represent the growing understanding of writing in the field; and affecting teaching and learning positively (instead of negatively) through assessment—and I show that as validity has evolved, newer definitions have addressed more and more of their concerns. Furthermore, as I detail a current convergence on using assessment to better serve teaching and learning in the concerns of both writing teachers and educational measurement specialists, I hope to demonstrate my agreement with Huot (2002) that writing teachers benefit by acknowledging and working with the contributions of educational measurement instead of against or separately from them, and vice versa.

Before moving on to do this work, however, I want to first acknowledge again the very real social, political, and theoretical obstacles that stood in the way of the evolution of validity and the meeting of teachers' concerns. Huot (2002) discussed a “theoretical entrenchment of many in the educational measurement field” (p. 43) which has ignored the contributions of English teachers. But more importantly, he showed that this “entrenchment” meant some in educational measurement held on to old conceptions well after changes had been accepted. Huot cited P. L. Cooper (1984) as an example since P. L. Cooper characterized indirect assessments as acceptable measurements of writing skills long after indirect assessments had been problematized by research studies of writing on such things as the writing process.

Related to what Huot (2002) noted, we see additionally that P. L. Cooper (1984) disregarded construct validity (which was well-established by 1984) in favor of old assumptions inherent in classical test theory. As discussed earlier, Williamson (1993)

showed that the origins of psychometrics in behaviorist principles and the search for abstract truth made it difficult, at first, for objectivity not to have primary weight, and his observation is seconded by Elliot (2005). As long as testing clung to objectivity, it could not make a turn toward recognizing the theoretical understanding of reading and writing in the field of composition which held that reading and writing were both processes involving the construction of meaning. The behaviorist and objectivist values in early psychometrics are exactly those principles that constitute P. L. Cooper's "entrenchment" and that contribute to a resistance even to the changing definition of validity expressed within educational measurement.

Additional obstacles are discussed by Elliot (2005) who showed the commitment of testers to assessments that were lower in cost and that could be administered to larger numbers of students. Elliot presented data on the increasing population of students entering higher education from 1879 to 1989, and the large population increases he outlined gives the reasoning behind testers commitment to efficiency and low cost—there were exponentially larger and larger numbers of students coming into higher education. These commitments to choosing assessments based on a serious consideration of cost and efficiency as primary concerns made it difficult to meet the demand for direct writing assessment called for by English teachers, at least not without developing some way to regulate and streamline the process. Elliot suggested, as well, that earning and profit played a role at times in motivating the kinds and types of assessments testers developed when he provided figures for the College Board's earnings during World War II. Earning far more from work for the government on large-scale tests for soldiers than from work for colleges and universities, the College Board (and later ETS) would, of course, feel

compelled to put their energy, resources, and research effort into large-scale tests and solving reliability issues over meeting the concerns of teachers and addressing validity.

Indirect Assessment vs. Direct—Reliability vs. Validity

The first concern expressed by English teachers historically—that writing assessment should involve the direct assessment of writing—was present even in the ETS documents that Elliot (2005) reviewed to compile his history. He referred multiple times to early key figures in testing who summed up (and at times expressed frustration with) the call from English teachers for tests that involved actual writing. Additionally, the call for direct assessments by English teachers was documented throughout the work of Edward White (1973, 1993, 1994, 2001). The concern of teachers had a lot to do with how firmly placed indirect assessment was as the primary writing assessment method prior to the 1960s. Since the focus of testers around this time was on an understanding of validity that accented criterion and content measures and that required a sufficient level of reliability in order to establish validity, as discussed earlier, indirect assessments were more congruent with these understandings.

Further, indirect assessments were popular because, as Camp (1993) argued, testing specialists lacked knowledge on the processes of reading and writing. Camp claimed that prior to composition's emergence as a field, the lack of knowledge that could inform the creation of better assessments stemmed from a shortage of studies on writing by English content specialists, and that once the field of writing developed, so did assessments. However, Huot (2002) suggested a different reading of Camp's explanation. Focusing on Camp's own lack of familiarity with those writing in the content literature, the overall struggle for composition as a field to establish its legitimacy, and the unequal

power position of teachers in discussions of writing assessment, Huot argued that even when studies existed, both sides remained ignorant of contributions of the other.

Although I would side with Huot here, regardless of which way the situation is read, it is clear that the lack of knowledge on the theoretical basis of writing *at use* in assessment creation meant that testing specialists found it acceptable to use error identification and correction as the primary assessment methods (Thorndike, 1911), and later, to use multiple choice tasks (Palmer, 1961). These indirect tasks more easily achieved acceptable levels of both validity and reliability as classically defined in psychometrics as opposed to direct measures of writing ability.

The focus on reliability over validity by ETS and testers frustrated English teachers precisely because it led to a rigid focus on indirect measures of writing like grammar tests. It was largely accepted to assess students writing through indirect means since attempts to assess students' writing directly had not yet yielded reliable results—i.e., raters could not agree at a level acceptable statistically for the College Board (Elliot, 2005; Greenberg, 1992; Huot 1990a). In a testing situation dominated by testers concerns, the worry over establishing acceptable levels of reliability in order to then secure validity (as well as concerns with cost and efficiency) continually trumped questions raised by English teachers. This was true even though developing changes to validity actually lined up well with English teachers' concerns about how it was possible for an indirect assessment to be seen as valid when it aligned with other indirect measures but did not include the actual task of writing and instead involved some smaller portion of the content of writing.

The dominance of indirect measures of writing would start to change when Diederich, French, and Carlton (1961) set out to examine reader disagreements and, in the process, determined that many of the factors of writing that readers mentioned correlated to each other. Godshalk, Swineford, and Coffman (1966) would go on to use this finding to develop holistic scoring—where readers express their impression of a whole piece of writing instead of on individual traits scored together. Through holistic scoring, Godshalk, Swineford, and Coffman secured acceptable levels of reliability. Despite this seeming victory, indirect measures remained the more popular approach to assessing writing well into the 1970s (White, 2001).

Concern About the Construct

The second concern of English teachers continually expressed in the literature pertains to the conception of writing at work in writing assessments (C. R. Cooper, 1977; Faigley, Cherry, Jolliffe, & Skinner, 1985; Gere, 1980; Odell, 1981; White, 1993). The field of composition, which by the 1970s and 1980s had established itself enough to have a substantial body of literature on the processes of reading and writing, clearly saw that the focus on grammar and correction inherent in indirect measures could never gauge fully all that was involved in writing ability. As a result, English teachers found little satisfaction in writing assessment measures that did not recognize and represent the growing complexity of writing they themselves could see.

Before I discuss an example from White (1993) of English teachers' dissatisfaction with the construct of writing represented in indirect assessments, I want to first discuss an earlier instance of construct problems that is related to, although not about, writing ability directly. The example illustrates a problem with the construct from

which testers operated even at the beginning of the field of testing. As far back as the 1874 entrance exam at Harvard crafted by Adams Sherman Hill, there was a concern with making sure students were prepared for the work they would find in college (Elliot, 2005). However, rather than worry about the correct class placement, the focus then was more on a students' fitness for college work. Early worry about the possibility of allowing underprepared or undeserving students admittance to college actually created the first movements of writing assessment. Despite the expressed interest in admitting by merit, it seems that cultural biases played a role (Elliot, 2005). But additionally, so did an overly simplified conception of literacy.

Elliot (2005) discussed men, such as Robert M. Yerkes (head of the American Psychological Association) and Carl Campbell Brigham (psychologist and field tester under Yerkes), who were instrumental in forming the growing science of psychometrics through their involvement with the first mass testing of intelligence for placement of army soldiers in 1918. It was from this work for the army that the first intelligence tests were accepted. But what this work for the army reveals is that these men, products of their time, saw no problem with their own biases against immigrants and black people. They developed a test that measured intelligence primarily through literacy. But instead of understanding literacy as socially and culturally learned, which would make illiteracy or dialect differences in literacy a potential barrier to a true assessment of intelligence, they equated literacy and intelligence. Unable to see the problem with their conceptions of literacy (and by extension their conception of intelligence), they went on to uphold their work and results as data in the field of eugenics, claiming proof of the inferiority of black people and non-Nordic immigrants (Elliot, 2005).

This problem in the conception of intelligence evidenced by the earliest testers is not so far off from the problems in the conception of writing made by those much later. White (1993) discussed specifically the problems with the ETS created Test of Standard Written English (TSWE) which he explained was designed to decide which students required remedial courses. However, as he made clear, the TSWE decided course placement through a multiple-choice test on Standard English. White (1993) argued that a large part of the problem with the TSWE was that it accepted “one of the most elitist aspects of American society” which was (and perhaps in ways still is) that “scholastic aptitude involves being born into an upper-middle-class white family” (p. 85). But he states additionally that the test

also accepts as fact that an ability to identify correctness of dialect is an acceptable definition of writing ability. While the ETS is well aware that editing is not the same as writing, those defending the TSWE point to high score correlations (among middle-class white students, of course) between editing and writing tests. Thus, there is no need to spend money to ask students to write, or to pay faculty to read that writing; by giving scores based on the family dialect of students, obtained cheaply, ETS helps colleges identify those who require remedial help in “writing.” (White, 1993, p. 85)

Therefore, bias—this time toward the relationship of writing ability to standard written English—played a role in the conception of writing at work in the TSWE. White’s concerns about how testers defined the construct of writing in their test represent another example of the serious problem with a definition of validity that does not include some discussion and justification of the assumptions and interpretations behind decisions.

Another nuance of the problematic understanding of validity at work in White's (1993) example (and the intelligence test example) can be found in Camp (1993). She explained the thinking of testers, specifically on multiple-choice writing tests, and her words reveal how an understanding of validity which focused mostly on criterion and content validity was at work in these indirect assessments. Camp (1993) explained that multiple choice tests were seen as valid because the tests covered some of the "skills necessary to writing" in addition to having determined "correlations between test scores" and other measures such as "course grades" (p. 47). These are exactly the procedures for establishing validity under the criterion and content models—figuring out how the measure correlates with other measures and how the test task represents the actual task. However, this explanation should also remind us of the weaknesses of these measures—weaknesses that construct validity sought to correct. The first weakness is the strong drive for testers to confirm their own judgment that the test task is useful for assessing the actual ability. The second weakness relates to the nature of writing ability as not fully observable, but theoretical. Given that arguments were already being made for construct validity in educational measurement (Cronbach, 1971; Cronbach & Meehl, 1955), a different procedure for establishing validity for the construct of writing was not only necessary, but paralleled exactly teachers' concerns.

Some explanation on the nature of the terms observable and theoretical in regards to writing assessment may be in order here. When writing is assessed, there is an observable performance (i.e., the student's piece of writing is produced), but English teachers, as they make sense of that performance, also operate with theories related to the understanding of reading and writing at work inside the mind of the writer. They may

employ, for instance, theories on how social factors affect students' reading and writing, on the processes (of cognition, decoding, composing, etc.) students go through when reading and writing, on how students' own conceptions of reading and writing change as they learn and revise, and so on. Because writing ability is both observable and not, it would make sense that, during this era (1970-1985), the measure of that ability would be validated through construct validity. Regarding construct validity, Kane (2001) stated, "The theoretical constructs and the indices used to measure them are validated by examining how well the theory as a whole accounts for the observable phenomena" (p. 335). Thus, as the definition of validity had already changed to include procedures for validating theoretical constructs, English teachers' concerns about the constructs at work in indirect assessments should have been addressed. This is particularly true given Kane's (2001) explanation that the theory behind a construct would need to be checked if it was disputed in a particular context (p. 334), as was the case with the disputed construct in writing assessment.

Why were the theories at work in indirect assessments not discussed or checked? Well, again, the idea that reliability was a necessary condition for validity was a factor. While holistic scoring had established acceptable levels of reliability, these levels were not as high as those for indirect assessments. Therefore, teachers would need to begin fighting for direct writing assessments and assessments that more closely represented writing as it was understood by the field. It took English teachers pushing for and working on direct measures like holistic scoring themselves before changes in the dominant forms of assessment were seen (White, 1993, 1994, 2001).

The most powerful influence affecting change to the dominant form of assessment though had to do with the conversation of writing assessment as it had already been established among testers and English teachers. Despite construct validity's inclusion of procedures for the validation of theories, (an inclusion that could have been fruitful for English teachers' arguments about the construct of writing represented in indirect assessments), Huot (2002) argued that the division present between testers and English teachers had already set the terms of future engagement. The conversation of testers concerned reliability, and English teachers believed they could succeed at securing direct assessment if they could convince testers of its reliability. Additionally, English teachers, who were not studying the literature of educational measurement on validity and who operated with a working definition of validity that centered on the loose expression that the test should measure what it says it measures, saw securing direct assessment through holistic scoring as a success on the front of validity as well (Huot, 2002). Thus, they concentrated on the adoption of holistic scoring (White, 1973) and saw it as a success because it allowed for the direct assessment of actual pieces of writing which, on its face, was directly measuring writing with writing.

But what English teachers accepted as a success could potentially be seen another way. Holistic scoring, as Huot (2002) reminded us, was developed through ETS as they worked to create direct writing assessment that could meet statistical requirements for reliability (p. 36). He argued that the impetus behind holistic scoring was acceptable levels of reliability (achieved by Godshalk, Swineford, & Coffman, 1966) and not validity. Further, this goal greatly impacted the view and treatment of writing inside the use of holistic scoring. Huot made this claim despite an acknowledgement that the way

holistic scoring was used and described by English teachers such as White (1993) allowed for a discussion of conflicting views on students' writing whereas holistic scoring as used by ETS sought to eliminate disagreements in order to ensure interrater reliability. To Huot, these positive differences were a result of English teachers working to adapt and control holistic scoring to meet their goals as teachers while also claiming a success for the direct assessment of writing. However, he concluded that in the process of advocating for holistic scoring and operating from an incomplete understanding of validity, English teachers mistakenly held up holistic scoring as working toward validity as they overlooked its origins as a method for achieving reliability.

Concerns About Teaching and Learning

The third concern of English teachers—that assessment should, above all other concerns, promote and serve our teaching and learning goals—has always been a part of the work of writing assessment. Even the earlier enumerated concerns about indirect tests and the construct of writing have at their heart a required usefulness of assessment for teaching and learning. To make this particular concern visible, I focus here on the era where English teachers fought for and succeeded in securing direct assessments through holistic scoring. Primarily, my goal is to establish that their concerns about teaching and learning, like the concerns discussed earlier, could be (in fact, as I argue later, would be) addressed by an understanding of validity as an argument (Cronbach, 1988; Kane, 1992, 2006). While these teachers are, of course, not the first to express this concern, and as we see later, they are surely not the last, I accent this era purposely to set up a later discussion of how holistic scoring's design as a method to achieve reliability limited its ability to fully address certain teaching and learning related concerns.

To begin, we can clearly see a deep concern with the goals of teaching and learning during this era in the work of Edward White. In fact, his work in the 1970s and the early 1980s to secure direct writing assessment through holistic scoring hinged on a desire for assessment that served the work and goals of teachers. To illustrate what I mean, I turn to White's (1993) recounting of the push for direct assessments by English teachers (through the adoption of holistic scoring) in the 1970s and early 1980s, and the reasoning behind this push.

As White (1993) discussed this history, he focused on some of the problems with indirect tests. Specifically, he talked about the College Level Examination Program's General Examination in English (CLEP). White argued that the CLEP English test, a multiple choice test that students could take to essentially get out of the required college writing class, was a huge concern in terms of teaching and learning because it made poor assumptions about (and thus placed absurd expectations on) the content of these classes and what they were meant to do. He wrote:

Since it was designed to test for achievement of the goals of freshman English, it presented a definition of the foundation course for most English departments in the country. That definition turned out to be even worse than that given by the TSWE. If the TSWE presumed to place students into freshman English by examining their dialects and editing skills, the CLEP test declared it could recommend the award of college credits by testing the same things. English faculties looking at the test were thunderstruck: Here was freshman English defined without including either reading (aside from the reading of test questions) or writing. (White, 1993, p. 87)

White went on to explain how the test was defeated by the Freshman English Equivalency Exam in California (on which he worked) and the development of holistic scoring, but the issues he raised make it clear that the impetus for English teachers' promotion of holistic scoring assessments was a concern about teaching and learning. Students who had no knowledge of college level reading and writing (as it was seen by teachers) could potentially be exempted from the course on the basis of a test that not only measured something else entirely, but suggested—to those students who took it and to the colleges who may accept the CLEP as credit for the course—that what was required of college level reading and writing was knowledge of the standard dialect.

Even when direct assessment was achieved through White's work in the 1980s (see White, 1993 further for a discussion of how quickly holistic scoring was adopted in a usually resistant field), White still felt the need to drive home a point about the relationship of testing to teaching. For example, as he described the "pitfalls" of testing in writing, White (1986) began by describing the conflict between testers and teachers and by expressing the value of having teachers work in assessment at all levels because they understand "the intimate relationship of testing to teaching" (p. 53). And, as he outlined how to create solid writing programs, most specifically as he discussed the issues of admissions, placement, and instructional evaluation, White (1989) made sure to discuss the conflicting goals of administrators and teachers and the need for bringing these goals into alignment (pp. 113-114).

Interestingly, the English teachers' concerns about developing assessments that served teaching and learning goals, as expressed in much of White's work, were compatible with developments in validity that were taking place around the same time.

Cronbach (1988) put forth an expression of validity that, through its wording, accented the idea that validity was an argument that would need to be made to different stakeholders: “Validation speaks to a diverse and potentially critical audience; therefore, *the argument must link concepts, evidence, social and personal consequences, and values*” (p. 4, emphasis in original). In fact, this expression makes it clear that there is an obligation to address those with opposing concerns or issues. Additionally, the newer understanding of validity includes requirements that the use of a measurement should have a positive impact on teaching and learning (Cronbach, 1989; Messick, 1989a, 1989b).

Centering on Validity

I have shown writing teachers’ concerns about indirect assessments, the construct of writing they represent, and the teaching and learning goals that are served by them, and I have discussed how these concerns actually align with developments in the definition of validity (while accenting specifically the time period of the rise of holistic scoring). Going forward, I want to discuss how the compatibility between teachers’ concerns and newer definitions of validity went largely unrealized until holistic scoring (and its accent on reliability) was questioned. By discussing this questioning, I hope to show that more recent scholarship, which takes the concern of teaching and learning as a central standard, does so precisely because English teacher-scholars began to move past a reliance on outside authorities (on classical test theory with an emphasis on reliability) to set the terms of useful writing assessment. Instead, teacher-scholars began to actively engage with emerging educational measurement literature in order to make use of that

literature (and its changing understanding of validity) for the achievement of their own instructional assessment goals.

As I have outlined above through a discussion of Edward White's work, holistic scoring was largely accepted as a dominant assessment method. But holistic scoring, with its history as a method to secure reliability (Huot, 1990a), was not without problems, particularly for teaching and learning. The focus on reliability in assessments had produced a concern with rater agreement in holistic scoring. Holistic scoring (as first developed by Godshalk, Swineford, & Coffman, 1966), required that raters be trained to achieve agreement in order to ensure acceptable reliability rates. This training to achieve agreement meant that samples, guidelines, or descriptions of writing would emerge representing the agreed upon ranking of writing. Alternately, any observation not pertinent to the scoring guidelines or to the achievement of interrater agreement was immaterial (in fact, as Myers 1980 described it, raters who disagreed or made immaterial observations would need to be retrained or removed by group leaders).

Perhaps the limits to teaching and learning inherent in the design of holistic scoring went largely unrealized by English teachers partly because, despite the flaw of its reliability-focused design, holistic scoring as adopted and used by English teachers seemed to stress agreement less and discussion more, and was, therefore, more faithfully committed to teaching and learning. For White (1993), while the end goal may have been to reach an agreement, the discussion of disagreements or different viewpoints was not barred. White discusses at length the valuable teaching and learning centered discussions that happened in holistic sessions and how they engaged teachers and created a sense of community. Therefore, even though holistic scoring was born out of concerns for

achieving reliability, it is important to acknowledge that many teachers saw and used holistic scoring as a form of assessment to serve teaching goals. Further, it also provided a way for English teachers to secure direct assessments. As a result, it served an important first step toward addressing some of English teachers' expressed concerns about assessments. But it was not long until the literature on holistic scoring began to acknowledge some of its limits and problems.

Several concerns emerged in the literature specifically about the process for securing agreement (and therefore, reliability) in holistic scoring. First, there was a concern that the achievement of agreement highlighted those superficial aspects of writing that raters could most easily agree upon such as length or handwriting (Charney, 1984; Huot, 1990b). Additionally, there was concern that the act of rating, with its focus on agreement, stifled the authentic reading of raters (Charney, 1984; Gere, 1980; Stock & Robinson, 1987) as it interrupted the natural process of reading. Finally, there was a worry that requiring agreement blocked an investigation of rater disagreements and the aspects of writing that become visible in those disagreements (Barritt, Stock, & Clark, 1986; Hake, 1986). This was particularly important given that the discussion of these disagreements had rich teaching-related value (Stock & Robinson, 1987).

These concerns were increasingly linked to holistic scoring's design for achieving reliability at the expense of other things. Reviewing the literature on holistic scoring, Huot (1990a) noted that the "very emphasis on reliability, which made holistic scoring procedures acceptable in the first place, has retarded their scope and stunted their growth" (p 204). He further argued that it was important to begin to replace reliability with validity as the chief concern. With this suggestion, Huot opened new avenues for his own

(and, of course, others') research. Shortly after examining the literature of educational measurement on validity, Huot (1996) summarized what he believed to be a positive development in validity theory:

The theoretical basis of educational assessment has been undergoing a radical series of changes. Lee J. Cronbach and Samuel Messick, two of the most important validity scholars in testing theory, have been revising their theories for the last decade....For both of them, the validity of a test is now inextricably bound to its use and effects. In other words, a test that has a negative influence on teaching and learning can no longer be considered valid, a far cry from traditional views of validity that center on the degree to which a test measures what it purports to measure. (p. 114)

To Huot, those working on validity in educational measurement had begun to recognize what many English teachers had long argued—that validity could not be attributed to a test if the test negatively impacted the teaching and learning of writing. Huot (1996/2009) expressed the change in educational measurement another way: validity theorists had begun to acknowledge the “social construction of knowledge” (p. 161). To Huot, this change aligned testing theory with writing theory and provided writing teachers with a way to make testing theory useful. Arguing that a theory of writing assessment was needed since, for too long, practice had occupied our work, Huot (1996) claimed that we should “align ourselves” with the changing “theoretical basis of educational assessment” because it “can strengthen our positions with testers and our ability to devise new and more appropriate measures for the evaluation of literate activity” (p. 114).

Since Huot (1996) expressed a need for writing assessment to develop theory to guide and support its work, he began that work in his landmark (1996/2009) essay “Toward a New Theory of Writing Assessment” by looking at and accenting writing assessment literature that did not take reliability (or classical test theory overall) as the central concern of assessment work. Specifically, he discussed Smith (1993); Haswell and Wyche-Smith (1994); Allen (1995); and Durst, Roemer, and Shultz (1994) at length. Smith (1993), working at the University of Pittsburgh, developed a system of placement essay reading which used course teachers to make decisions on placement for their course of expertise. At Washington State University, Haswell and Wyche-Smith (1994) also relied on more expert teacher-raters, but only for the difficult decisions determined after a first wave of reading by less experienced teachers had placed the majority of students. Allen (1995) found that portfolios from a variety of colleges and universities could be shared and read online by outside, expert teacher-readers for the purpose of program assessment so long as the directions and guidelines for the portfolio were made clear. Durst, Roemer, and Shultz (1994) of the University of Cincinnati focused on an exit exam reading of portfolios with groups of three teacher-readers who openly discussed their views at length to make a decision about a students’ exit, but also to help develop teachers and the program.

Huot (1996/2009) found commonalities across these pieces, even as the specifics differed, since, as he stressed, “The focus of each of these programs is inward toward the needs of students, teachers, and programs rather than outward toward standardized norms or generalizable criteria” (p. 165). Looking at what theoretical assumptions underpin all of these works, Huot (1996/2009) outlined five principles for writing assessment:

assessment should be “site-based,” “locally-controlled,” “context-sensitive,” “rhetorically-based,” and “accessible” (p. 171). O’Neill, Moore, and Huot (2009) added “theoretically consistent” to this list (p. 57). These principles were clearly reflected in the Conference on College Composition and Communication’s (CCCC) *Writing Assessment: A Position Statement* (2009) which expressed that assessment should be developed with local, contextual, and pedagogical goals in mind along with current theories on learning, writing, and assessment.

I have spent a significant portion of this chapter building a reading of writing assessment literature that has culminated in the expression of principles that arise largely from Huot’s work. Huot’s body of work built on a dismissal of reliability as a key focus for writing assessment—a term he claimed was problematic because of its lineage from the psychometric tradition. Yet, he embraced validity, at least newer definitions, as key for writing assessment work going forward. Lynne (2004) problematized a reliance on both of these terms. She argued that the “active support” of “psychometric principles” as “appropriate for writing assessment” by scholars such as Huot and White “authorizes supervision of composition’s procedures by the educational measurement community, but also limits our ability to theorize our own practices” (Lynne, 2004, p. 7). There is some wise caution to Lynne’s critique. Just as Huot, looking to the past, found reliability to be incongruous with the concerns of English teachers, as we advance our knowledge in the field of writing and assessment, it may be that the term of validity ceases to serve our purposes, or like reliability, it may create a set of problems that can only be shed by a future shift in emphasis or terminology.

However, I disagree with Lynne's depiction of Huot's logic as "[limiting] our ability to theorize our own practices" (2004, p. 7). The theoretical basis that emerged from Huot's (1996/2009) work comes not from the terms in educational measurement, but from a close examination of writing assessments developed outside of those terms (Allen, 1995; Durst, Roemer, & Shultz, 1994; Haswell & Wyche-Smith, 1994; Smith, 1993). Thus, Huot's theory was not "supervised" by the terms of educational measurement, as Lynne (2004) suggested, but served by it. This is supported as well by my earlier argument that validity has long been developing in ways that align it with teachers concerns which suggests not a secondary status for writing teachers, but instead specific ways that educational measurement has had to catch up to the knowledge of English teachers.

Given the pieces Huot (1996/2009) examined and the principles of writing assessment theory he gathered from the work of writing teacher-scholars, as well as my own reading of the assessment literature, I would argue that validity provides writing teachers with a way to make testing theory useful for the communication of their arguments. In fact, Huot (2002) claimed that the rhetorical training of teachers makes them well-adept at crafting arguments, such as the validity arguments that would need to address various stakeholders. Given that writing teachers are not the only party interested in the writing ability of college students, I believe Huot's use of validity serves us well in making a case to other constituents.

In summary, while the importance of teaching and learning to the work of testing, as we have seen, has long been a concern of teachers, more recent scholarship has taken teaching and learning as a central concern or standard for assessment going forward. This

focus has come about because of an acknowledgement that measures of writing were too narrowly concerned with achieving interrater reliability when the value and usefulness of those measures was in question. By questioning this focus and instead accenting questions of validity (as expressed by educational measurement in terms useful to our work), writing assessment scholarship has moved to determine the theoretical principles to guide us going forward.

William L. Smith: Validity, Expertise, and Context

While English teachers were calling for assessments that directly assessed writing, that represented the reality of what we know about writing, and that positively influenced how writing was taught and learned, and while educational measurement theorists were beginning to push against the idea that validity resided in a test, we have seen that it was some time before everyone was on the same page. Additionally, I have shown that when the right conditions began to emerge, a number of scholars—both content specialists and educational measurement theorists—began to ask important questions about the focus for writing assessment and its adherence to psychometric terms and the “epistemological foundations of positivism” (Moss, 1992, p. 233) that did not seem to serve our needs.

Going forward, I want to look at William L. Smith’s (1992/2009, 1993) studies investigating writing placement at the University of Pittsburgh. Responding to a concern with reliability, but not entirely focused on it, Smith (1993) was interested in the disagreements of his raters, so he examined those disagreements closely, and he realized (1992/2009) that experts (those who had most recently taught the class for which they were deciding placement) seemed to have fewer disagreements when paired with other

similar experts and when making a decision involving the particular class they had taught.³ My focus here on Smith (1992/2009, 1993) is central to the present validity study at Cardinal College since the college follows a modified version of Smith's expert reader placement model. An understanding of the structure of Smith's model and what he found is helpful for upcoming work in Chapter Three where I describe Cardinal College's system as I sketch the interpretive argument on which it rests. Therefore, I begin this section by detailing Smith's studies and his findings.

Additionally, a close review of Smith's (1992/2009, 1993) work is necessary because it exposes the importance of expertise, context, and validation as essential tenants for future assessment work. Therefore, after I detail his studies and findings, I make two key points: First, while Smith made no direct mention of educational measurement literature and chose not to use the term validity, he did discuss the adequacy of placement decisions. Despite his avoidance of the term validity, Smith's concerns with the effectiveness of his placement system, the clear expression of the ideas that guide it, and the continual study of that system are closely related to the concept of validity that began to emerge from validity theorists around the time of his studies. Thus, this similarity reveals again that writing teachers' concerns align well with validity and the requirements of validation that stem from it. Second, while Smith was not the only (or first) scholar to assert an importance of expertise or context (see for instance, Haswell & Wyche-Smith, 1994; Pula & Huot, 1993; Stock & Robinson, 1987), his work is unique in that his findings largely put to rest his (and our) concerns with reliability and they directly influenced current writing assessment theory. Smith (1992/2009, 1993) found

³ It is important to clarify here that most of the studies discussed in Smith (1993) actually precede the published findings of Smith (1992/2009). Smith (1993) discusses several studies that he conducted into his program that were, at first, unpublished. This is why I discuss Smith (1993) first.

that a teacher's expertise in the context of a certain course (and not strict methods for achieving agreement) allowed expert teachers to achieve higher levels of agreement since he discovered that disagreements on certain students' placements were actually agreement that those students exhibited characteristics outside of or across multiple courses. Further, their expertise in specific course contexts allowed placement readers to make valuable judgments and decisions directly that aligned with teachers' perceptions of students' placement in the courses and the students' course grades (Smith 1992/2009, 1993). With the value that Smith attributed to expertise in context, it is no surprise then that the current theory of writing assessment accents context and teacher expertise as central principles (CCCC, 2009; Huot, 1996/2009; O'Neill, Moore, & Huot, 2009).

Detailing Smith's Studies

To unpack each piece in more detail, let me first begin with Smith's "Assessing the Reliability and Adequacy of Using Holistic Scoring of Essays as a College Composition Placement Program Technique" (1993) in which Smith summed up eight years of research into the University of Pittsburgh's placement test on such subjects as "the reliability of raters, the decisions made by split resolvers, the adequacy of the raters' judgments, and the training of raters" (p. 142). Prior to detailing his specific studies, Smith (1993) offered a brief review of literature on direct assessments, a description of the courses and testing program at Pitt, and a discussion of general "problems specific to placement testing" (p. 150). A few of the problems he discussed are key to details of his studies.

First, Smith (1993) discussed the problem of assumptions about the scale for placement rating. Smith (1993) said those of us studying and deciding placement assume

that each scale point in placement has a range of the same size, we assume that “raters have similar ranges for each [scale] point,” and we assume that “the scale point [ranges] about” (p. 152). He showed that it is theoretically possible for serious disagreements about placement to arise when the reality does not meet assumptions. Another problem has to do with the “fit” of students to the scale points (and therefore to the courses they represent) (Smith, 1993, p. 155), which can again create disagreements. Disagreements that arose from the problems Smith detailed were significant because they were not a problem of rater reliability.

Smith (1993), interested in examining the possibility for non-rater related disagreements, discussed a first brief study ruling out the possibility that raters might have “a distribution mind-set” (p. 154) such that they would rate essays to create the expected distribution of students across courses. He selected previously rated essays to create a set that violated the usual distribution and had raters rate them as if new. He found that Pitt raters did not have such a mindset. Feeling assured that raters would not be swayed by such expectations, he began looking more closely at the reliability of raters.

Studies of reliability. Smith (1993) discussed a number of studies that dealt specifically with reliability. To set up the issue of reliability, Smith (1993) made a distinction between “disagreements about placement” and “disagreements about quality” (p. 160) for as he explained, readers will usually have different opinions on quality, particularly if they are approaching the student’s work as literature (as in the Pitt program), but a disagreement on a student’s placement is not the same. A placement disagreement may indicate a problem with the raters’ consistency, but it could also be attributed to a problem with “scale” (Smith, 1993, p. 160). In the case of placement, the

scale is derived from the courses (of which there were three course options—Course A, Course B, and Course C—and the fourth additional option of exempting a student from the program), and, as Smith (1993) explained, disagreements could mean that students “do not fit neatly” into the courses as designed (p.161).

Having well established a potential problem with scale point ranges that differ significantly between raters and the possibility for gaps in these ranges that do not neatly place students in a course, Smith (1993) first sought to study how Pitt raters resolved disagreements. Discussing the use of split-resolvers—where an essay receiving two different ratings is read by a third rater or more until agreement is reached—Smith outlined how it was customary for Pitt raters to put essays that split initial raters into a separate stack. Therefore, the third (or subsequent) readers knew they were resolving splits. Smith was concerned that these raters, knowing that they were split-resolvers, would not read and make decisions as they would otherwise. Following a modified procedure where split essays were fed back into the initial stack until an agreement was reached, Smith secured a first resolution or decision on placement. Then, he placed these essays into a split-resolution stack to get a decision from a rater who knew they were determining placement for a disputed essay. Comparing the results on each split essay, Smith (1993) found that the raters who knew they were split-resolvers only agreed with the unknowing raters 69% of the time, which suggested that the reading was not the same when a rater knew they were resolving a split (p. 161).

While split resolving was an ongoing interest for Smith (1993), he studied it as part of an ongoing concern with rater reliability. Therefore, after mentioning it briefly, he turned to the program’s interrater reliability, which had been calculated for some years as

ranging between 45% and 70% (Smith, 1993, p. 161). Since he proposed that rater disagreements could “be a product of natural disagreements about students who do not fit neatly into courses” (p. 161) rather than of poor rater reliability, Smith (1993) turned to study intrarater reliability (raters’ agreement with themselves). Such a concern, he theorized, could remove an issue with differences between raters and more clearly reveal an issue with the match of students to the courses. Smith designed a study that slipped essays that were read and scored in a first session back into stacks of essays read for sessions three and five to investigate raters’ consistency with their own past decisions. He found that raters were fairly consistent, rating 75.7% of these essays the same in sessions one and three, 64.3% of them the same in sessions one and five, and 75.7% the same in sessions three and five (Smith, 1993, pp. 163-164).

While these numbers were better than the interrater reliability rates for these same sessions, which varied across the sessions from 40.3 to 64.2% (Smith, 1993, p. 161), the disagreement even between separate readings by the same rater encouraged Smith to look further. Reanalyzing the numbers, he found that intrarater “agreement increas[ed] with each course level” (Smith, 1993, p. 164), and that if an essay had received a split decision across the first two raters, then a single rater was more likely to show inconsistency when reading the essay again in later sessions. That is, Smith found that the *disagreements* about placement (across and within raters) revealed an *agreement* that a particular student was not a clear fit for available courses.

Since the essays that provoked splits could reveal this underlying agreement, Smith (1993) proposed that determining the “rater-set reliability”—the reliability across two sets of raters—may be a more worthwhile measure since it did not mask the

important information about a mismatch between students and courses (p. 166). In subsequent studies designed specifically around rater-sets, Smith set out to compare two sets of rater pairs and found that these rater-sets agreed highly. If the first raters produced no split, the second set agreed 93% of the time (Smith, 1993, p. 169). If the first rater-set had produced a split of one scale point difference, the second set recreated that split 71% of the time (Smith, 1993, p. 169). The other portion of the time, the second set produced no split at all but still selected the rating (with one instance of exception) that was ultimately decided on by split-resolver of the first set (Smith, 1993, pp. 170-171). Finally, when the first set of raters produced a split that spanned two scale points, the second set returned the same result 65% of the time (Smith, 1993, p. 171). The remaining percentage of the essays produced a one scale point difference. The second set of raters “never produced a nonsplit” on these essays (Smith, 1993, p. 171). Smith (1993) explained that “These results indicate that the placement ratings given by these trained, qualified raters are reliable even when the raters disagree and, thus, appear to be unreliable” (p. 173).

Smith (1993) discussed several other studies of reliability, the details of which, while interesting, are not as pertinent to my study. I mention them though because the conclusions Smith reached in these studies support earlier important assertions. For instance, Smith mentioned additional investigations into split-resolvers which indicate further that split-resolvers read differently when they know that split resolving is what they are doing. Also, Smith (1993) discussed data on a third set of raters from outside the University who agreed with the initial raters at a high level overall (72%) but with significantly decreased agreement on splits that widened as the splits became larger (p.

171), indicating further that expertise in the context of the program was an important aspect of placement rating.

Before moving on to discuss studies on the adequacy of placement, I discuss the final studies of reliability that Smith reported on, which were recounted in both Smith (1993) and Smith's "The Importance of Teacher Knowledge in College Composition Placement Testing" (1992/2009). Given his earlier findings about rater agreement, even on splits, Smith decided to investigate further the splits of more than one scale point, as these showed the least amount of agreement. Smith (1992/2009) looked into the backgrounds of the raters in his pairs, and he found, as he reanalyzed the initial data, that the course the rater had most recently taught had an effect on their decisions since "almost all of the disagreements...happened when the raters had most recently taught different courses" (p. 175). Evaluating data collected from think-aloud protocols, he found additionally, that when the raters were reading an essay for a student who they decided belonged in the course they had most recently taught, they talked about more than just the student's essay—they imagined how the student would be in the class. From there he reanalyzed splits from prior years and proposed that teachers with recent teaching experience in a particular course have *privileged course knowledge* about their students that was far more useful than any training done during placement (Smith, 1992/2009, p. 183).

Seeking to further confirm his theory, Smith (1992/2009, 1993) designed a study which selected essays previously read by rater-sets with six different combinations of course teaching experience (some pairs shared experience and others did not), and for a new placement session, he assembled new rater-sets in pairs with identical course

teaching experience as the initial readers. Once the pairs were set up with identically experienced initial raters, Smith (1993) fed the old essays randomly into stacks of new placement readings. He found that when the new sets of raters matched in experience to the original and were reading nonsplit essays, they agreed 100% of the time with the original raters as opposed to an agreement of 81.7% for rater-sets with unmatched experience (p. 180). Further, when the matched rater-sets were reading splits, they agreed 83% of the time as opposed to the 16.7% agreement for rater-sets that did not match the experience of the original set (Smith, 1993, p. 180). He concluded that “all raters, regardless of course taught expertise, are able to reliably discern students who are prototypic of a course” (Smith, 1993, p. 181), but the same was not true for other students.

Looking closely at the decisions being made for specific essays by raters with various combinations of course teaching experiences, Smith (1992/2009) focused further on when the raters were making decisions “accepting” or “rejecting” students from the course with which they had recent experience (p. 184). He found that decisions were even more reliable when a particular rater or rater-set was making a decision specifically about the course they had most recently taught. From there, Smith (1992/2009) proposed a placement reading procedure where expert readers would use their “Most Recent Course Taught expertise” to make a decision regarding placement into their course or rejection from it, either higher or lower (p. 183). In the process of using expert readers, Smith ultimately solved issues with reliability, since his expert readers actually agreed at a higher level than holistic scorers.

Studies of adequacy. Since establishing a rater's reliability, as Smith (1993) said, "still does not mean that [raters'] decisions are correct or that they make the best decisions for the students" (p. 185), Smith held that adequacy would need to be determined as well. However, Smith noted that there were difficulties with checking adequacy directly or during placement because ethically students could not be misplaced and the data which was needed to check placement naturally occurred after the initial decision was made. This, he stressed, meant that there must typically be a retest during the first weeks of the course. Additionally, he discussed, in two different sections of his piece, the data sources for checking placement, including first-week retests, teachers' perceptions of correct placement, students' perceptions of the same, final grades, and exit exams/posttests (Smith, 1993, p. 156).

Detailing the flaws with these data sources, Smith (1993) outlined his decision to focus on teachers' perception of adequate placement primarily and final grades secondarily. He noted problems specific to each of these data sources that have affected his study design. For instance, Smith (1993) claimed that the time frame when teachers' perceptions are most helpful is after some knowledge of the particular students is built up (three weeks in) but before students start making gains within the class (after week five) (pp. 158; 186). And the calculation of final grades Smith (1993) used was the median grade (and if a student was above or below the median) since final grades vary by teacher and often take into account student issues or attendance factors that are not about the adequacy of placement. (p. 186).

Once he outlined these issues, Smith (1993) started to connect the adequacy data he collected as part of the earlier reliability studies to the course teaching expertise

findings. He had teachers of Course B and Course C rate their students on a scale that included describing students as “prototypic of the course,” “slightly above,” “slightly below,” “could have been placed in lower course” and “could have been placed in higher course,” and since there was nothing lower than a Course A placement, these teachers used the same scale but with the below and lower options removed (Smith, 1993, p. 188). The results indicated that less than 8% of students were considered inappropriately placed, but given the different distributions for the course with 14% of Course A rated as too high and 8% of Course C rated as too low, there seemed “too few” students being placed into Course B (Smith, 1993, pp. 187-188). Reanalyzing the data to separate out scenarios where the decision of placement into a certain course involved at least one rater that had teaching experience for that course (and later when both raters had such experience), Smith (1993) found that “the probability that the student [was] misplaced [dropped] to near zero” (p. 188) when an expert was making the decision.

Smith (1993) went on to examine the data for certain patterns of raters and ratings he was interested in. Specifically, he looked at ratings with a variety of decisions (accept, reject, or split) regarding Course B and Course C. These decisions were made by rater-sets with shared Course B versus shared Course C experience, as well as mixed Course B and C experience (with one rater on either side). He wanted to see if such data confirmed his earlier finding that teachers’ perceptions would vary in relation to the course expertise of the rater-set. He found that the teachers’ perceptions varied in each instance just as his theory predicted, and the final grade analysis confirmed the findings as well. Overall, he concluded that the data confirmed the adequacy of the raters’ decisions.

The comparison of teachers' perceptions of students in Course B who had been placed by two Course B raters versus the perception of teachers on those students placed by two Course C raters specifically confirmed that the expert decisions resulted in less of a sense that the students were misplaced (Smith, 1993). Smith (1993) concluded that the very low level of misplacements that existed overall were a product of "the wrong raters making the judgments" (p. 192). Upon making this discovery, Smith (1993) set up a method of rating around course teaching expertise where "each rater's decision was limited to 'accept' or 'reject-high' or 'reject-low'" and "each essay was read by at least two raters until it was 'located'" as being in Course A, B, or C; exempt; or between one of these four possible placements (pp. 192-193). Then, taking the decision on the placement of those students between placement options, what Smith (1993) termed *tweeners*, out of the hands of raters—since past studies showed that they fared well grade wise in higher placements—Smith set a policy that those between courses go into the higher placement option, except for those between Course C and exemption (p. 193).

Subsequent studies and discussions, after the rating method changed, confirmed that distributions and teacher perception data changed in ways congruent with expectations (Smith, 1993). Further, Smith (1993) found that most raters felt more at ease and confident with their decisions in this method, and, despite a system that could potentially (and did) increase the number of readings raters had to do, the amount of time for the rating procedure overall was not increased as anticipated but was actually quicker. Smith attributed this to the fact that raters had to make fewer decisions in the new method and the decisions they did make were served by their expertise. Additionally, an analysis of rater agreement showed that "Course A raters agreed on 96% of the essays they read

together, Course B raters agreed on 93%, and Course C agreed on 86%” (Smith, 1993, p. 196). Smith (1993) attributed the drop in Course C raters’ agreement to fatigue from their involvement in reading 90% of the total number of essays (p. 196). In the end, Smith asserted the usefulness of his findings but called for more study into the method of using raters’ teaching expertise since his studies only addressed the application of placement and there were additional concerns to address.

Smith and Validity

Throughout this chapter, as I have discussed the history of writing assessment, I have shown that it was common for discussions of direct writing assessment to focus on the achievement of reliability. I have argued that this focus was due in large part to the relationship of reliability to validity in early testing theory—i.e., reliability was a prerequisite for validity. The relationship of reliability to validity coupled with a desire to secure direct assessments of writing caused even content specialists to give great attention to the issue of reliability. Brian Huot and Michael Neal (2006) named the focus on reliability in writing assessment a technological focus that solves reliability issues at the expense of validity (pp. 424-426). Interestingly, they offered Smith’s (1992/2009, 1993) work as one way to reframe this problem. Smith, as Huot and Neal (2006) said, was interested not in reliability alone, but in “the accuracy and adequacy of the decision being made for placement” (p. 426). Huot and Neal pointed out that Smith’s concerns and solutions—interested in the placement decision being made—actually aligned more with a newer definition of validity emerging around the time of Smith’s inquiry (Messick, 1989a) as opposed to those studies overly concerned with reliability. In this section, I consider this claim using some of the details of Smith’s (1992/2009, 1993) studies in

order to outline my agreement with Huot and Neal (2006) (later Huot, 2007 and Huot & O'Neill, 2009) that Smith's work does show a connection to newer understandings of validity.

Before I go on to detail my agreement that Smith's (1992/2009, 1993) studies relate to newer concerns of validity, I want to first clarify that Smith never directly expressed these newer understandings of validity or cited the work of validity theorists. In fact, he actually avoided engaging with the literature on validity and, to an extent, the term validity as well. This avoidance of validity theory is seen in the brief mention he made of validity as he began to discuss what research he felt was needed on direct assessments:

Little research has been conducted on holistic rating for placement testing, and the question of validity has not been sufficiently addressed. Because validity carries a considerable amount of baggage..., I, instead address the question of "adequacy of placement" into writing courses. This avoids controversy about the kinds of validity and allows me to focus more precisely on what I consider to be the real validity issue for placement testing in writing. (Smith, 1993, pp. 143-144)

As we see here, Smith set out not to consult or use definitions of validity in his work, and such a choice was a conscious one. Smith purposely avoided the term validity—choosing instead to focus on the concept of adequacy—because while he acknowledged that it was validity that must be established, he wished to keep himself from becoming embroiled in a debate over the types of validity. A debate which he felt would not allow him to get to what, for him, was the essential question of validity in placement: Is the student well placed?

Smith's (1992/2009, 1993) work is held up as representing a change from the tight focus on reliability in the writing assessment literature (Huot, 2007; Huot & Neal, 2006; Huot & O'Neill, 2009) because, while a concern with reliability motivated his first questions about the placement program at Pitt, Smith (1993) explained from the beginning that he was much more concerned with the adequacy of placement decisions (p. 143). Therefore, his concern, regardless of his avoiding the term validity, was an emergent concern of contemporary validity theory.

Smith's (1992/2009, 1993) work aligns well with specific principles of newer theories of validity. As I discussed at the beginning of this chapter, more recent theory conceives of validity as an ongoing process where testers are asked to question their assumptions; to look at the decisions made in testing and their consequences; to outline an argument for their program using current theoretical knowledge and clear evidence from ongoing study; and to consider how to convince various stakeholders. Going forward, I offer a few examples taken from Smith's work that illustrate his concern with these very principles.

Validity requires ongoing research that questions assumptions. Shortly after naming that studies of adequacy (validity) were needed in writing assessment research (placement specifically), Smith (1993) discussed why program administrators tend not to see adequacy of placement as an issue. He revealed that, as a field, we do not question the assumptions on which our beliefs rest. Talking specifically about how placement programs tend to think their placements are adequate since few students are moved during the first week of class, Smith discussed his own investigation into the reasons

teachers of particular courses at the University of Pittsburgh gave for not moving students. Having asked teachers to share their thinking, he found,

Their responses indicated that they do not want to make many changes on their class roster; in particular, they don't want new students coming into their courses after the first or second week of class. Thus, they do not move as many students as an external reader might, yet external readers are seldom used because the expense is too great. Therefore, our sense that our method is adequate is based on only a belief, a belief which, not surprisingly, keeps us from questioning the placements and the methods. (Smith, 1993, p. 144)

Here Smith demonstrates a commitment to asking questions even about things that are usually overlooked. Smith thought about common assumptions—in this case that the small numbers of changed placements suggests that the placement was adequate—and he considered that there may be other explanations that were not being seen or considered. As a result, he goes on to study those other explanations by investigating his program closely. In fact, the details of all his studies, as I have described them above, show a strong commitment to continual inquiry that searches for evidence instead of relying on assumptions and comfortable beliefs, a core facet of validation as I've outlined it in this chapter (Cronbach, 1989; Kane, 2006; Messick 1989b).

Validity requires a look at decisions and their consequences. There are other points of connection between Smith's work and our understanding of validity. As Messick (1989a) made clear, those involved in assessment must think about the decisions that are made using a test. Because Smith (1992/2009, 1993) centered his work on the decisions made, and his findings were about who was best able to make those decisions,

he argued that teachers are experts at placement decisions as they have current experience and knowledge of the course they are teaching, which contextualizes their reading of student work in ways not typical of holistic raters. Further, Messick (1989a) expressed that the consequences of a decision should play a role in a determination of its validity, and that emphasis on consequences is perfectly echoed in Smith's (1993) rationale for why the decision (and its adequacy) is a key focus for placement:

What makes placement testing so different from other types of testing is the direct impact on each student. Any mistakes have a personal effect on the number of courses the student must take (and possibly on the tuition bill) and even on the student's sense of self-worth. A student who believes she is a competent writer but who is required to take a "remedial" course (in some colleges, such courses carry no credit) is likely to feel both frustration and self-doubt, and the parents who are paying for the courses are distressed. (p. 159)

Smith clearly talks about the consequences of placement decisions here as he mentioned the number of courses that will be taken by students, the cost, and even the impact on the student's "self-worth," all of which show the lasting effects of poor placement decisions. Regardless of Smith's avoidance of the term validity, his concern with ensuring appropriate placements and his extension of that concern to the consequences of inappropriate or, as we will see momentarily, unclear placements is a concern with validity as we have come to understand it.

Validity, as an argument, must be convincing to various stakeholders. Smith (1993) continues to exhibit principles of the current understanding of validity in his discussion of how the potential consequences of placement (and the real people they

impact) necessitate explaining decisions to others (Kane, 1992, 2006). As he continued to discuss the potentially misplaced and frustrated students and their parents that he mentioned in the passage above, he stated,

My experience with such students and parents has taught me that public accountability is a necessity. If I can provide data on our placement testing, data to which appropriate administrators are privy, students and parents are much more likely to accept the placement decision. Moreover, students who accept the decision are less likely to resist the teachers' efforts. (Smith, 1993, p. 159)

As we clearly see here, other people besides writing teachers are interested in, and therefore must understand, placement decisions. The impact of inappropriate placement on students, both financially and emotionally, was something for which Smith felt he must account. He wanted to ensure that others could see the program was making efforts to keep misplacement from happening. Further, Smith was also aware that he must convince students, parents, and administrators of the soundness of the reasoning behind placement decisions even when he found that students were not inappropriately placed. The decisions must be made clear; otherwise, a student's resistance to the placement may have the same effect as that student being placed in an inappropriate course.

Validity requires an argument using current theory and evidence. Convincing various stakeholders of the soundness of decisions, as we have seen above, requires data that they find persuasive. Since validity is “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness of inferences* and *actions* based on test scores or other modes of assessment” (Messick, 1989b, p. 13, emphasis in original), to argue that Smith

(1992/2009, 1993) understood this principle, we must see Smith using what he knows about current theories of writing, reading, and learning and the data he collected from his ongoing study of the program to make an argument for the assumptions and the decisions of his program. Smith began both reports on his studies (1992/2009, 1993) with a detailed description of the courses in the program and the theories of reading, writing, and learning on which they rest. Additionally, he continually collected data on the decisions of his raters—one study led him to new questions, concerns, and subsequent studies. And, he continually questioned the assumptions of the program, as we saw earlier, but he even questioned the assumptions of the field of writing assessment when he asked if focusing on the reliability of raters would hide potential problems with students who do not fit into the courses. Doing all this, Smith's work comes quite near Messick's (1989b) definition of validity.

I have insisted previously that the understanding of validity currently at work in educational measurement is useful for meeting our concerns as teachers and (in agreement with Huot, 1990b) for shaping the arguments we make to colleagues, educational measurement specialists, administrators, and various other stakeholders. Smith's (1992/2009, 1993) work makes the usefulness of validity visible since it reveals both what he achieved in terms of his own program's understanding (and the field's), as well as what he was able to show others about the adequacy of the program.

Smith and Expertise in Context

In addition to revealing the importance of validity and the usefulness of validation, Smith's (1992/2009, 1993) studies have another value. Asking a question about adequacy of placement, Smith naturally became far more interested in what

adequacy meant inside his local context over any generalizability of his findings. Yet, despite the situated nature of his inquiry, his findings do expose a principle which is useful across writing assessment contexts: direct writing assessments benefit from readers with contextualized teaching expertise related to the goals and purposes of the assessment. Such readers can make adequate decisions because they have expertise in specific teaching and learning contexts. In sum, Smith's findings reveal the contextual nature of knowledge (and of writing programs), and his work allows for teacher knowledge and our particular contexts to become core considerations over the usual considerations of large-scale assessments such as controlling for agreement and generalizing across contexts.

To clarify how the principle of expertise in context is important for writing assessment, I begin by discussing two examples. The first shows that Smith's (1992/2009, 1993) expert reader—a reader who has course teaching experience and is making decisions regarding the course with which they have recent teaching experience—eradicates the need for the kind of rigid reading procedures and alignment to a shared rubric of the past. The second demonstrates how a focus on specific teaching environments and the value of expert readers contextualizes what is required in prompt development. After discussing these two examples, I go on to expand my argument for contextualized expertise to include how it is developed within an interpretive community that is actively studying and discussing disagreements.

Expertise in context vs. agreement and generalizability. I have already discussed in several sections how the focus in previous writing assessment literature was on reliability, specifically interrater reliability. With such a focus, holistic scoring and its

specific methods and procedures were devised to secure sufficient levels of agreement between raters (see for instance, Davis, Scriven, & Thomas, 1981; Myers, 1980; White, 1994 for descriptions of procedures on achieving agreement). Myers (1980), for example, devoted a whole chapter to scoring (pp. 30-47), and a large part of that discussion talked about how one ensures alignment. From the selection of anchor papers to how to handle raters who consistently score papers out of line with their fellow raters, the focus was on making sure there was agreement.

More than just showing an example of procedures for securing agreement though, Myers (1980) is also interesting to consider as his description of how teachers were used to develop rating criteria presents an important contrast to the use of expertise Smith (1992/2009, 1993) detailed. When rating criteria were developed for a particular holistic scoring session, Myers (1980) described how the choosing of anchor papers and the development of a rubric began with experienced raters who were classroom teachers (p. 1). To arrive at the criteria that later raters would be trained to agree upon, the teachers who were most experienced with rating, who later became “table leaders,” selected essays from the stack that were prototypical of score categories (Myers, 1980, p. 33). Once they had these “anchor papers,” table leaders recorded “the characteristics” that “differentiat[ed]” each score (Myers, 1980, p. 33). Then, from the descriptions and judgments of these experienced raters (who, remember, Myers said should be teachers), the criteria were developed.

While Myers (1980) stated that teaching experience was important, in this instance, it was not the teaching experience that was used in the assessment going forward. Rather, the focus on agreement and generalizability meant that the initial

criteria, criteria which may have been influenced by teachers' expertise inside specific teaching contexts, were analyzed outside of any acknowledgement of that expertise or its context. Instead, the context of the scoring session and the need for agreement take precedence. The experience of teaching was only necessary, it seems, for show since this experience was not tapped in the scoring or used to make decisions useful for teaching and learning. For these raters (both the table leaders and general readers), teaching experience was, in effect, replaced by the agreed-upon criteria that represented the definition of quality writing which raters were trained to uphold.

In short, Myers (1980) seemingly acknowledged the importance of expertise, but that teaching expertise was rendered impotent by holistic scoring procedures which achieved agreement by ignoring writing and teaching characteristics that emerged in specific writing and educational contexts. This holistic scoring procedure stands in direct contrast to assumptions behind Smith's (1992/2009, 1993) work. Instead of assuming that readers must be made to agree, and that the only useful way to get meaning out of that agreement is to have criteria that are context free and generalizable, Smith trusted in the experiences his teachers brought to their placement reading, and he considered that the teaching context not only played a role, but was essential to make useful placement decisions. Indeed, his findings supported that the teaching context did play a role.

Prompts in context. The literature of holistic scoring procedures that focuses on securing agreement is not the only set of literature that must be reimaged to make use of contextualized expertise. Prior to the 1980s, the discussion of developing writing tasks and prompts in relation to assessments was not focused on particular contexts. Many writing assessment studies proceeded as if all prompts—even when asking for

rhetorically different kinds of writing—were equal in assessments (see Ruth & Murphy, 1988 for a discussion of this literature). Then, studies began to emerge that did not treat prompts as if they were generally equivalent. Hoetker (1982) reviewed some of that emerging literature on essay topics which were just beginning to look at the effect of “mode of discourse” or “rhetorical context” on the writing produced (pp. 382-386).

From this literature review, Hoetker (1982) stressed how little was known and how much more study would need to be conducted. He specifically called for studies into how students understand writing prompts since, as he claimed, “we know very little, beyond the anecdotal and suppositional, about how students actually read the texts of essay topics” (Hoetker, 1982, p. 389). Murphy and Ruth (1993) noted a similar lack in literature looking at how students understand writing prompts. Interestingly, they connected this lack directly to the importance of context as they described a common practice of borrowing prompts created for other assessments which often had very different purposes or were used for placement within very different curriculums. Using Messick’s (1989b) definition of validity, Murphy and Ruth (1993) argued that writing prompts were not valid in and of themselves but only in the context of their use. They demonstrated that a test must be validated inside a contextualized use, and they discussed how the psychometric system of field testing prompts, with its emphasis on statistics and expert judgments of prompts, limited what could be seen about the differences among students’, prompt creators’, and essay test readers’ understandings of a topic.

I do not disagree with either Hoetker (1982) or Murphy and Ruth (1993) that it is necessary and important to know how students understand writing prompts; however, I want to look more closely at specific pieces of what both Hoetker and Murphy and Ruth

say since I believe they present an interesting contrast to each other. Despite their agreement that we know very little about students' reading of prompts, Hoetker and Murphy and Ruth approach the value of students' reading of prompts from differing orientations. Hoetker (1982) distinguished assessment done for research purposes from that done for teaching purposes, and he claims that assessments done for research would need to follow different methods and procedures, particularly for scoring essays, which would better lend themselves to the statistical differentiations that would be necessary to show clear differences in the effects of specific topics. While he explained well why he believes that such a distinction between teaching and research would help us arrive at clearer findings on the differences between writing topics, his claim rests solely on a need for statistical differentiation. In the case of writing topics, which represent a construct of writing, statistical differences in how students perform do not seem to be the only, or even the most helpful, information required. Statistical differences between the performances of students on writing prompts will not provide enough information for teaching and learning decisions. Further, I wonder about the practical value of research findings if they follow procedures so different from assessments for teaching and learning.

The problem with relying on statistics is something Murphy and Ruth (1993), in contrast to Hoetker (1982), made clear. Murphy and Ruth questioned specifically how a reliance on statistics will show how topics are interpreted. What Murphy and Ruth (1993) say reiterates my concern about the value of research that does not add to what we know in ways valuable for teaching and learning:

The conventional procedures in field tryouts produce massed data which are insufficiently revealing of individual processes of interpretation. Statistics may tell us *when* some effects are occurring but not *what causes* them. It is not enough to know that some topics induce better scores than others; we need to know why. (p. 284, emphasis in original)

Murphy and Ruth argued we need to know why and how students interpret prompts. Without this information, they questioned what the value of statistically differentiating between prompts would be. I believe the how and why of students' interpretations is absolutely necessary for teaching and learning purposes. The statistical differentiation Hoetker desired in research would only matter if it helped us to apply that knowledge to teaching and learning, and while it may offer some information, it does not offer the most useful information for teaching and learning, as Murphy and Ruth showed. Therefore, assessments for research cannot really operate so separately from those for teaching purposes. Instead, assessments, when researched, must be studied in particular contexts of teaching and learning, as Murphy and Ruth recommended at the close of their piece.

While Murphy and Ruth (1993) presented a helpful critique of the usual procedures for the field testing of prompts, while I believe their call for qualitative data is essential to the study of writing assessment in general, and while they forefronted and accentuated the importance of context, I want to look closely at their arguments through the lens of Smith's (1992/2009, 1993) findings. I believe that the importance of a teacher's contextualized reading expertise, demonstrated by Smith, offers a way to tease out moments where Murphy and Ruth (1993) describe students' interpretations of prompts so that the value for teaching and learning is revealed. To begin, I want to look

at how Murphy and Ruth detailed the reasoning behind their argument for qualitative data by first discussing typical field testing.

Describing the two usual ways writing topics are field tested by psychometricians, Murphy and Ruth (1993) say that first, psychometricians consult experts on the “fit between particular test items and specific descriptions of the content domain to be assessed by those items” (p. 270). Second, psychometricians conduct field trials using prompts to statistically compare student performances to determine how difficult the question is and how well it differentiates “between strong and weak students” (Murphy & Ruth, 1993, p. 271). Murphy and Ruth (1993) claimed that both of these procedures “aim to detect and reduce sources of error and to improve the items for their intended purposes” (p. 270). They questioned aspects of a psychometric and statistical orientation as they suggested the use of qualitative data, and called for “a model of writing assessment that respects the legitimacy of alternative interpretations of the writing task” (Murphy & Ruth, 1993, p. 268). I want to use Murphy and Ruth (1993) to call into question the assumption that differences in interpretation are “sources of error” in order to see the potential in these differences for improving teaching and learning (p. 270).

To illustrate exactly how variations in students’ interpretations can be useful for teaching and learning, I want to look at a moment in Murphy and Ruth (1993) where they summarized earlier work which noted differences in students’ prompt interpretations. In this summary, they described information that, given what I have outlined from Smith (1992/2009, 1993) about how expert readers make decisions, should stand out as useful data for expert teacher-readers in particular contexts (and additionally should be

interesting phenomenon of study to learn more about the development of discourse expertise in students). Here is how they described interpretation differences:

During the 1980s, we reported results of investigations which suggest that readers of the topic (both the student writers and the teacher raters) choose among cues embedded in the text of the task, both honoring and ignoring elements which may enable them, with varying degrees of success, to match the test maker's intentions and expectations (Ruth & Murphy, 1988). (Murphy & Ruth, 1993, p. 268)

Murphy and Ruth claimed that the interpretation of a prompt involves recognizing "cues" that are in the prompt, and that to do well on a writing task, students would have to understand those cues in ways similar to the expectations of the "test maker" and/or test readers. Instead of seeing different interpretations as errors, as in psychometrics, Murphy and Ruth (1993) called for us to respect "unexpected but plausible constructions of meaning" (p. 297). Connecting their argument to Smith, we can deduce that expert teacher-readers would best be able to distinguish the *plausible* interpretations from other interpretations on writing tasks inside their particular contexts.

Let me explain. Take for instance a program, such as Cardinal College's, which integrates academic reading and writing. The placement writing prompt and the prompts for checking placement administered on the first day of class are written to signal clearly to those familiar with the discourse of academic reading and writing that they must analyze one given reading passage using the information and ideas in another.

Differences in interpretations of the prompt, made visible in the students' written responses, are, therefore, useful to expert teacher-readers as they provide potential insight into students' previous academic reading and writing experiences. When the students

misunderstand what they should do, what they do instead reveals their thinking (and incidentally their own expertise). In short, I think that what Smith (1992/2009, 1993) showed us about placement readers—i.e. that their disagreements are sometimes due to the differing expertise of some readers—is applicable here to the students as well. Some readings are expert, and others may be developmentally related to that expertise.

For example, in the context of Cardinal College's prompt, it is common for students to write a response that merely summarizes each reading passage separately rather than writing the intended analysis. As a teacher in this program, such a response reveals to me that these students' lack familiarity with the signals for, or the writing of, academic analysis, as well as a corresponding familiarity with summary. My understanding of such a response has been gained from seeing many students in my courses with similar writing habits who, as we work in class, produce similar interpretations of course prompts that need teased out and problematized. The misinterpretation of the prompt will not dictate alone if I come to believe the student belongs in my particular course or not (I would take into account a number of additional things such as how well the student does with the task of summarizing that the student has set for him or herself, the student's use of attribution, any potential overreliance on the words or sentence structure of the text, etc.), but such information is essential to forming an understanding of the student's skills and needs inside the context of my classroom (a course which has specific objectives that may or may not meet this particular student's needs).

It is my belief that expert teacher placement readers at Cardinal College have insights into the students in their courses which help them recognize similar patterns in

responses during placement. Thus, as Murphy and Ruth (1993) called for qualitative study into the interpretations the students themselves describe and a healthy respect for variance in interpretation, I feel assured that this type of work and the orientation of respect could lead us to study, analyze, and theorize the thinking behind our students' interpretations in ways useful for decisions about our teaching and their learning.

Recognizing the usefulness of variance only arises when we research real contexts, and when we recognize the role expertise plays in teaching and learning based determinations.

Now that I have discussed some of the literature of prompts, it is important to outline the links between this literature and my earlier discussion of scoring agreement. The problem in both of the above scenarios—scoring agreement and writing prompts—is captured well by Huot (2002) who wrote, “Conventional writing assessment’s emphasis on uniformity and test-type conditions are a product of a testing theory that assumes that individual matters of context and rhetoric are factors to be overcome” (p. 85). Thus, we see the influence of a testing theory that wished to find some way around what were, and are, essential concerns for the task of writing—rhetoric and context. Yet, there is no way for readers to come to agree outside a particular writing (and reading) context. Smith’s (1992/2009, 1993) expert reader model involved purposeful reading that stayed true to the importance of context and rhetorical knowledge in writing theory. Purposeful reading—the reading professionals do every day—does not require an additional external rubric or alignment because that work is the contextualized work experts are already doing. Further, since expert teachers have expertise inside a curriculum, they require, not a general prompt, but one which can best expose the information necessary to make an

expert decision in a particular context. There is no prompt that would work equally in all contexts or be interpreted identically by all students and teachers. Therefore, we need to study and specify our particular contexts while making use of the contextualized expertise of teachers.

Improving teaching and learning through the development of contextualized expertise. Smith's (1992/2009, 1993) studies implied that the disagreements of the past that alignment, training, and rubrics were trying to eradicate, happen less inside an expert reading, where the readers are a part of what Fish (1980) called an *interpretive community* (p. 14). I want to consider further the idea that the disagreements that do occur inside an expert reading (Barritt, Stock, & Clark, 1986; Hake, 1986; Stock & Robinson, 1987), much like Smith found, would reveal interesting information that has the potential for improving teaching and learning. In particular, I want to think about how a discussion and consideration of disagreements can actually be used to develop expertise inside an interpretive community.

Some disagreements, as Smith (1992/2009, 1993) has shown, arise not from raters' inability to agree on criteria but from their difficulty deciding what to do with students who do not fit neatly into the categories dictated by the courses. When a student does not fit into our categories, this may have a lot to do with what our categories do not encompass. If we consider this through the view of teaching and placement, Smith (1993) suggested that perhaps a student does not fit because our current curriculum does not address all of that student's needs. The disagreement, then, is essential to considering what to do with that particular student in terms of placement. Where should the student be placed? Are there more students like this one? Are there enough of these students to

warrant changing the courses? If not, what existing course is the best placement? These are crucial questions for determining the adequacy of the curriculum and the adequacy of placement, but more importantly, they are teaching and learning questions.

Disagreements happen when two raters, reading within the natural meaning-making process of readers, see something different in the same piece (Barritt, Stock, & Clark, 1986; Stock & Robinson, 1987). Going forward, I want to discuss the work of Barritt, Stock and Clark (1986) and the work Stock and Robinson (1987) since they, like Smith (1992/2009, 1993), investigated disagreements. Unlike Smith, however, these pieces were not primarily interested in validating placement decisions, but instead they investigated what can be learned from disagreements. I hope to show that these studies, taken together with Tierney and Pearson's (1983) "composing model of reading" (p. 1), can expose what it takes to develop contextualized expertise and how that expertise can have an impact on teaching and learning.

Barritt, Stock, and Clark (1986), working as part of the University of Michigan's English Composition Board in the reading of placement essays, discovered that despite their clearly set criteria and many years reading together as Fish's (1980) interpretive community, they could not eradicate disagreements. Thus, they set out to investigate their disagreements. They had readers actively discuss the pieces that received the most disagreements, in particular the criteria that led to the decision to place the student where they did. From there, they focused on the "discernible sources of disagreement" (Barritt, Stock, & Clark, 1986, p. 316), and they found that the placement readers discussed more than just students' writing; they also discussed "the imagined student writer" and "the writer as prospective student—as someone who could be helped in one or another

instructional setting” (Barritt, Stock & Clark, 1986, p. 319)—a finding not unlike Smith’s (1992/2009) observation that expert readers could talk about the writer as a student in their classes.

Thus, Barritt, Stock, and Clark (1986) concluded that their disagreements exposed something potentially important about the natural process of reading because “interpretive habits of mind become most apparent” through disagreements (p. 323). But, I feel they also revealed the importance of the reading process in the context of a teaching and learning environment when they wrote,

To require evaluators to pay attention only to pre-established criteria is to imply that this is the only way for fair readers to act. Such a stance makes reading very difficult for anyone who wants to take seriously her own response as an expert, particularly as an expert teacher, while at the same time trying to reach consensus with other readers. (Barritt, Stock, & Clark, 1986, p. 324)

A reader who is forced to restrict their reading to only what can be agreed upon cannot, as they made clear, examine their own expertise as it relates to other readers. And, this is particularly important for teaching related expertise because teachers must examine what they know and how they know it to develop and refine their expertise as part of an educational program of study.

The work of Tierney and Pearson (1983)—who argued that just like writing, reading is about the active construction of meaning—is helpful here in teasing out the development of reading expertise and, by extension, the development of reading expertise in the context of teaching. Tierney and Pearson detailed ways that readers make decisions about how to read, about their purpose in reading, and about the stance they take in

relation to a text, all of which will affect the meaning they arrive at from a text. Readers, in their description, must have experiences with various texts, as well as various purposes for, habits of, and ways of reading, within some context in order to develop the useful reading knowledge they will draw on when approaching future texts. Given Tierney and Pearson's work on detailing the particular stances and approaches readers can take, the way readers alter their plan in response to the text as it unfolds before them, and how readers revise their understanding as they go, the expertise of teachers is clearly related.

Teachers are expert at reading student work and taking the appropriate stances that allow them to see the contextualized pedagogical value of a student's text. And, they develop this expertise within the context of teaching in their locality, including time spent negotiating meaning with others in the teaching community with whom they work to carefully consider and make conscious which observations and interpretations influence their high-stakes decisions. In short, I am saying that expertise at teaching writing, of course, involves the ability to make meaning of students' texts inside a context, but a key part of making meaning is an ability to reconsider the appropriateness of an arrived at meaning given new or divergent information.

We can see some of these requirements for expertise in Smith (1992/2009, 1993). When Smith used expert readers in placement, he ensured that these readers recently taught the course about which they will make decisions, and that they taught the class for more than just a semester or two. Without some time teaching a course, a teacher has not yet seen enough in terms of specific students, texts, and scenarios to develop a sense of the students in that particular course. But, in addition to having experiences with the particular students within a course (and their writing), there are other requirements that

come about from my discussion of Tierney and Pearson (1983) above. New teachers, or even teachers new to a particular course, benefit from a discussion of their reading of student work, including a discussion of disagreements, since such disagreements can be used to refine understandings of students' texts and explore different interpretations and ways of reading student work. Such a discussion of students' work (including the disagreements) by both experienced and new teachers, maintains an interpretive community that is working together to make explicit their shared goals and collective vision.

Going forward, I want to talk a little more about the value of interpretive communities for teaching and learning by discussing Stock and Robinson (1987) which detailed work on the Ann Arbor Assessment from 1984-85. To open their piece, Stock and Robinson (1987) discussed how the "gap between the teaching and the testing of writing" is caused by the different "needs" of "administrators and policy makers" versus "teachers and students" since these groups must do different things with the assessments (pp. 93-94). The need of administrators for general, broad numbers is a problem because "to reach such findings—to make their construction possible—testers must make assumptions that are antithetical to the best current theories of how written language is used and how its various uses are best acquired or learned" (Stock & Robinson, 1987, p. 94). These assumptions have to do with removing both the students' work and the criteria for assessing it from the context surrounding their creation. In short, they argued that local assessments created by teacher-researchers better meet the goals of teaching and learning than large scale assessments. This is an argument I have already discussed and supported in detail through an analysis of other pieces of the literature of writing

assessment. But what Stock and Robinson add to it is an acknowledgement of the contrasting emphases of large-scale assessments and local assessments that make local assessments so valuable for teaching and learning.

Nowhere is Stock and Robinson's (1987) argument for the teaching and learning value of local assessments clearer than when they discussed specific questions asked by teachers that make the abstract concerns of large-scale assessments concrete. For example, they outlined questions of task development:

Do I ask my students to write compositions similar to those my colleagues ask of their students? Together do we provide our students enough diversity in what we ask in our various classrooms to make their studies interesting without providing them so much diversity that they are confused about the purpose of English studies and the values English teachers hold as a community? (Stock & Robinson, 1987, p. 103)

Additionally, Stock and Robinson (1987) included questions on scoring:

Why are we awarding these compositions high scores? These, low ones? What are our values? Do we share values? Is *showing* valued over *telling* among English teachers? Would our colleagues who teach social studies find these abstractions in student writing that are unsupported by concrete exemplification as troublesome as we do? Can I change my evaluation of a composition to agree with my colleague's evaluation? Is her argument for her reading convincing to me? (pp. 103-104, emphasis in original)

These questions, and additional ones they asked, deal with such things as the relationship of the task to the curriculum and to individual teacher's expressions of the curriculum.

They spoke of the “diversity” of ways for teachers to enact the curriculum while still being consistent to departmental vision, to understandings of writing as a process, and to the needs of the larger college community. The goal of these questions is to make improvements to teaching and learning through assessment, but these are improvements inside a specific educational context and community. Outside of local assessment, such important questions cannot be asked; they are superseded by concerns with generalizability. Outside of local assessment, the reality that variation is involved in reading, writing, and teaching cannot be recognized; it is hidden behind the assumption that such things are error. Outside of local assessment, the contextual expertise necessary for negotiating and making meaning of variance is unacknowledged; it is misunderstood as content area knowledge rather than praxis.

By thinking about the specific contexts of our programs, their theoretical underpinnings, and the expertise of our teachers through a process of ongoing study, I believe writing programs, no matter what particular system of placement they use, can begin to make conscious their knowledge and expertise. This knowledge can be linked back to the specific learning goals and objectives of their local courses and programs, and further, it can be used to revise practices, courses, and goals as theoretical understandings, student populations, and teaching personnel change. With its forefronting of context and expertise, Smith’s work invites us to think about the local context and the decisions that could be made when writing assessment focuses its attention less on the technical matters of assessment and more on the decisions of *teachers* for the benefit of *learners*.

The Relevance of Validity, Expertise, and Context in Placement Today

The essence of placement testing is the collection of some kind of information or data to aid in making an appropriate decision about which course in a writing program would best match a particular student's needs. While it is easy to describe this core essence of placement testing, a lot of the details of placement testing vary and what has been seen as acceptable has changed dramatically across time. The kind of data collected and the method of interpreting it have changed and varied as placement testing has developed. Haswell (2005) offered a list of common placement data and methods:

1. Essay written by the student and read by English department teachers, sometimes by teachers across campus.
2. Essay written by the student and read by a teacher of the target course, who sometimes accepts or rejects the student into his or her course.
3. Essay written by the student and read and scored by a certifying agency or testing firm outside the university, the score then used by the university to place the student.
4. Essay written by the student and scored by computer software (e.g., Educational Testing Service's E-rater, ACT's e-Write, College Board's WritePlacer), which score is used to place the student.
5. Folder or portfolio of the student's high-school writing submitted and read by faculty to place.
6. Short-answer or bubblesheet test taken by the student, on which evidence of "verbal skill"—criterion-referenced score or norm-referenced percentile—the student is placed ("indirect testing").

7. Placement by the student herself or himself in the writing sequence, the decision based on information provided by the college, such as high-school GPA in English courses, scores on a writing or verbal examination, average success of test groups in various [sic] courses (“informed self-placement”).
8. Placement by the student himself or herself using the same kind of information but also relying on advice from counselors (“directed self-placement”).
9. Enrollment of all students in the regular writing course, but after a few weeks some students are placed into a more basic course or or [sic] given the added requirement of hours in a writing center, the decision made by their teacher or by a panel of faculty on evidence of course performance up to that point.
10. Delay of enrollment in writing courses, with possible requirement of a first-year writing course on recommendation from teachers of other courses taken during the first semester or first quarter. (para. 3).

Haswell named methods and practices that have been discussed within the scope of this chapter so far such as indirect and direct methods (including systems like Smith’s (1992/2009, 1993) expert reader which was characterized in number two), as well as newer practices I have not discussed in detail like computer scoring, portfolio assessment, and directed or informed self-placement.

Over the years, there has been an interest in figuring out what placement systems are most often used, but has been difficult to get institutions to report this information. Huot (1994) discovered this difficulty when he sent out surveys to 2,695 colleges and universities inquiring about their placement systems and 1,099 returned those surveys (p.

64). Given this sample, which Huot was careful to describe as similar to the population but underrepresentative of private and small institutions, Huot (1994) found that 51% reported that they used “a writing sample as at least part of their placement method” (p. 53). Yet 42% used ACT or SAT scores, tests not designed for measuring writing, and indirect tests of writing were frequent as well with 20% using the Test of Standard Written English (TSWE) specifically and a “preponderance” of the 26% who chose “other” indicating some multiple-choice grammar or usage test too (Huot, 1994, pp. 54-55). While no comprehensive counting has been published since Huot, Elliot et al. (2012) offered a few numbers for particular placement methods. They write, “29%...of Title IV postsecondary institutions used ACCUPLACER” and “directed self placement appears to be used by less than 1%...of postsecondary institutions” (Elliot et al., 2012, p. 287). And Hassel and Giordano (2011) noted that community colleges rely heavily on either or both ACCUPLACER (62%) and COMPASS (46%) (p. 30).

Given my assertions thus far though, in agreement with CCCC (2009) that placement testing should be connected to the local-context, and in agreement with Huot (1996/2009, 2002) that validation of our writing assessments (in line with newer definitions) is useful and necessary, the particular method used for placement is not as important as the principles and theory that guide the choice of that method for a specific local context. While the choice for placement should be a local one, that choice must be grounded in current theory; must align with local students, teachers, and curriculum; and must balance practical testing concerns. Going forward, I want to discuss some of the placement methods that have developed since Smith’s (1992/2009,1993) expert reader system. Since much of my argument so far has built up to a close reading of Smith’s

work which—despite my best efforts to make clear its relevance—may appear outdated given the time that has passed since the publication of his studies, I want to take a moment to offer a few brief examples of what I think may be at stake if we forget the three core principles I have discussed in this chapter that come from a close examination of Smith’s work: validation, expertise, and the local context. By situating this discussion in more current placement methods, I hope to show the continuing relevance of these principles in placement today; there are potentially great futures in newer developments for placement, but not if in using them, we forget the principles we have learned in research and study so far.

Problems with Validity

There are a number of studies that seem to discuss the validity of placement focused exclusively on the placement’s predictive validity. Such studies do not acknowledge the developments in validity theory which centered the focus of validity on construct validity and moved to see validity as an argument. Kane (2006) claimed that for placement studies, this focus on predictive validity is quite common. Yet, I wonder, given the real issues I have already outlined about how the construct of writing has been poorly represented by specific tests or measures historically, how can such studies ignore developments in the understanding of validity that require considerations of construct. Take, for instance, Mattern and Packman (2009) who completed a meta-analysis of 47 studies on the predictive validity of ACCUPLACER and never mentioned validity as an interpretive argument or the assumptions behind the test’s construct of writing. Or Scott-Clayton (2012) who acknowledged that validity is in the use of a particular test, but then focused her efforts on establishing the predictive validity of one institution’s use of

COMPASS so exclusively that, when she reached the discussion section of her study, she can hypothesize a disconnect between the courses and the test, but has no data on that issue to analyze. I think it is a serious issue when a study presents predictive validity and the computation of statistics alone as all that is needed to validate a placement test without a close examination of the construct of writing in the test and how that aligns with the goals of a particular writing program. It seems these studies, and others like them, are stuck in an old theory of validity. Even when they seem to say the right words, as Scott-Clayton does, the focus of their validation does not follow through.

I have worked at a community college where this is the case. The college uses ACCUPLACER's Reading Comprehension and Sentence Skills tests to decide placement into the developmental reading and the writing program. Outside of setting cut scores for these tests, there has been little examination of the content of the tests and how it aligns with the courses to begin validating the use of these tests at the college. Elliot et al. (2012) revealed one obstacle to validation as they discussed validating the use of ACCUPLACER at one university: they found that ACCUPLACER test makers will not publish the information necessary to study things like the difficulty level of specific test items (p. 296). While difficulties in getting information from test makers are real and important, I wonder if Elliot et al. skip over what seems like the larger issue: such tests have little to no theoretical support (Hassel & Giordano, 2011) given what we know about writing, and making a convincing argument for these tests will take far more than examining item difficulties.

Discounting Expertise

Interestingly, as Elliot et al. (2012) attempted to validate ACCUPLACER, they used the expertise of their teachers to get around their lack of data from test makers on test items: “Attempting a qualitative solution, we asked five experienced instructors to take ACCUPLACER three times—under the persona of an honors student, a traditional student, and a basic writing student” (p. 296). While they had their teachers pretend to take the test as if they were certain students—in essence, tapping the teachers’ experiences regarding what such students are like—they seem to see this act and the qualitative information it yields as a second-choice. The secondary status Elliot et al. assigned this information seems odd given what we know about teachers’ course expertise. If teachers can accurately take the tests as if they are students at various levels, why can Elliot et al. not skip the middle step here and ask their teachers to assess students’ performances? At times, some studies seem to reveal a basic distrust in letting expert teachers make decisions (and of course working to study and continuously cultivate that expertise).

An additional example of a placement method that has an odd relationship to expertise is directed self-placement (DSP) (Royer & Gilles, 1998). The theory that informs DSP was built on the cultivation of motivation and engagement in students. Therefore, the goal is not to supersede the students’ authority, but to help them take ownership of their learning and their decisions by empowering them to make the placement decision (Royer & Gilles, 1998, 2003). Despite an understanding of its goals, however, I, and the field, still have serious concerns. At its best, DSP provides students with important information such as class descriptions, examples of work, scores on

writing (or other) tests, and even the recommended placement of an advisor or teacher (Gere, Aull, Perales Escudero, Lancaster, & Lei, 2013). But, not all of this information is provided in every DSP program. Sometimes, the student can be given little more than course descriptions and a questionnaire about their past writing experiences (Gere et al., 2010; Harrington, 2005).

Even when a student is given significant information, I worry that in some contexts—contexts where students have little experience with the kind of reading and writing that is expected in the particular courses in the program—it may be quite a struggle for students to arrive at the intended interpretation or understanding of the courses in provided course descriptions. Students with little writing and reading experience similar to the writing and reading expected in a particular program may not really know how to read or understand the examples of work or course descriptions they are given. In short, I am saying that sometimes DSP presupposes that we have students who can already easily make sense of the kind of writing we will require of them. The ideas of DSP are attractive because teachers know that when students take ownership of their learning, it can improve that learning. However, depending on how different our way of teaching is from students' past experiences (which, with No Child Left Behind, may be a past dominated by standardization and national testing), we may be asking some students to make a decision about their needs when they have little context for understanding what their needs will be inside our curriculums. In fact, a large part of my own courses involves teaching students to *unlearn* since I ask students to critically interrogate their past learning about writing and their own past writing practices. In my context, where a significant portion of the population of students has reading and writing

experiences that differ dramatically from what they will learn in the course, I cannot see an effective use of DSP since it would be the equivalent of asking me to try to teach the student the whole course within a few hours and one interaction so that they can decide if such a course would be helpful for them.

Given my concerns, I am happy to see work that is seeking to answer Neal and Huot's (2003) call for "continued validity inquiry into DSP" (p. 245). Gere et al. (2010) and Gere et al. (2013) offered examples of ongoing work to validate DSP for a particular context. In fact, the findings of Gere et al. (2010)—that DSP at the University of Michigan does not produce a valid decision since the questions asked of students on the DSP questionnaire stressed length and correctness over the crafting of an argument which was more in line with the course—informed Gere et al.'s (2013) follow up. Gere et al. (2013) studied changes made to the questionnaire to address how well it represented the construct of writing at work in the courses and added a DSP essay. Work like this will be necessary going forward with DSP, and the field needs to see more on how the expertise of teachers and advisors, through the recommendations they give students, is used within the DSP program.

Problems with Context

Validity requires an acknowledgement of a particular context (Messick 1989b), and expertise only exists inside a context. Therefore, as I have discussed problems and concerns above, context has impacted each example. Since writing programs themselves differ in what they expect, require, and teach; the data, methods, scoring, and procedures of placement should best enact theoretical understandings inside the particular goals of the program and the constraints of the local context (constraints like cost and time). Take

for example, portfolio placement. On the level of expertise, a portfolio of work to look at when deciding a student's placement can mean additional writing samples on which to base interpretations of the student's readiness. Further, these samples are not only of timed or finished products, but a portfolio also includes drafts and reflections showing the process of a student's work across time. Even Hester, O'Neill, Neal, Edgington, and Huot (2007) have successfully extended Smith's (1992/2009, 1993) model of expert reading to portfolios.

In terms of validation, portfolios make sense theoretically for the same reasons—they line up well with writing theory, defining writing as a process and recognizing that writing ability varies by task and time. However, portfolios can be impractical inside some local contexts that have constraints of time and money; portfolios will take longer to read with their additional samples, so there will be a longer turnover time for scores, and by extension it will cost more money to pay potential readers. There is an additional concern about portfolios for placement in relation to particular student populations. Students who have done a good deal of writing in high school will have many pieces to choose from to assemble a portfolio, but not all students will have such writing experience. Further, given the opportunity to submit a portfolio of their work, many students will not have the inclination. That is perhaps why portfolio placements are optional like at Indiana University of Pennsylvania and at the University of Louisville (Hester et al., 2007).

Conclusion

Smith's (1992/2009, 1993) texts are landmark ones for writing assessment. Many writing assessment scholars see Smith's work as representing a key moment of change

shifting the focus of writing assessment from the reliability of assessments to their validity (Huot, 2002; Huot & Neal, 2006; O'Neill, 2011). Additionally, Smith's (1992/2009, 1993) work parallels a contemporaneous move in educational measurement redefining validity. Validity, in this redefinition, is no longer inherent in an assessment itself, but is an argument that must be created for the use of an assessment in a particular context (Cronbach 1971; Messick 1989a; 1989b; Kane, 2006).

By asking a question of adequacy and pursuing the systematic study of his own local program, Smith (1992/2009, 1993) offered an example of validation. Further, he demonstrated how one could study local, context-specific assessments, and his findings, which reveal the importance of teacher expertise in making placement decisions, showed the significance of such study. Smith's (1992/2009, 1993) findings revealed that teacher readers have complex knowledge and course expertise that they could use to make adequate decisions.

By conducting validation research and publishing it, we will come to understand more about the complicated interplay of our expertise as we develop it. Through this research, we can widen the literature of writing assessment to make it relevant to situated and specific contexts, student populations, curricular goals, etc., which may be important for helping us learn about others with similar contexts, but will be far more important for creating and promoting a culture of teaching and learning based assessment.

Teaching and learning based assessment is naturally contextual, local. It is a part of a writing program's particular philosophy and their construct of writing and reading. It is born out of our knowledge of the field, our situated positions, our student populations,

and the larger mission or goals of our institutions. Therefore, we need a better understanding of how the teachers within our contexts put their expertise to work.

As higher education faces the same assault of standardization that has stripped our middle and secondary school colleagues of their expertise and reduced the richness of their students' reading and writing, it is more important than ever to study our specific writing programs. Without validity inquiry into our own local programs and assessments and without the crafting of a cogent argument about their value, we will have little to combat a similar reduction in the opportunities to learn we must provide to students, most specifically basic writing students, at the college level.

An argument must be made for our use of assessment measures. We are not just meeting some requirement to assess that comes from outside or higher up. Instead, we are using our knowledge and expertise and acknowledging the reality of our local contexts. We are speaking about writing to those who do not have our expertise. We are describing the concerns of our context and encouraging others to avoid seeing our findings and decisions outside of their situated nature. We are measuring not just for the sake of ranking, but because we are focused on meeting the needs of students. And, we are pushing our expertise ever forward through reflective practice and the discussions and negotiations that must take place in assessment that is concerned primarily with teaching and learning.

CHAPTER THREE

METHODS AND PROCEDURES

Recounting a history of writing assessment that began with a divide between the work of testing and the work of teaching, Chapter Two built on Huot (1996) to argue that there are valuable understandings on both sides of the historic divide that are useful for writing assessment. Since, in the eyes of English teachers, the kinds of writing assessments historically advocated for by testing specialists (reliable but indirect tests) have often fallen short of representing the complexity of writing and addressing the concerns of teaching (C. R. Cooper, 1977; Faigley, Cherry, Jolliffe, & Skinner, 1985; Huot, 2002; White, 1973, 1993), such an argument may seem difficult to swallow. Yet, at the heart of my argument was an explicit discussion of how newer understandings—particularly of validity—in educational measurement have actually better aligned testing theory with writing theory and the teaching of writing.

Specifically, the stress in current validity theory is on crafting an argument for an assessment that includes an examination of the theory that supports the assumptions behind it, interpretations within it, and decisions made on the basis of the use of a test, and contains an examination of alternative interpretations and the consequences of a test's use (Cronbach, 1988; Kane, 2006; Messick, 1989b). This current theory better aligns testing theory with writing teacher's concerns about assessments that lack or have outdated support in writing theory and that have negative consequences for teaching and learning. In addition to stressing this alignment, my argument also built on the work of Smith (1992/2009, 1993) and forefronted the predominance of the authority and experience of writing teachers who, I argued, must be directly responsible and involved

in the validation of our own programs since our expertise gives us specialized knowledge that would improve how writing assessments address the goals of teaching and learning in particular programs.

If, as I argue, writing teachers are to be responsible for the validation of our own programs, we need, then, a stronger understanding of what it means to validate according to existing theories. In order to make convincing arguments to various stakeholders, we must know the language and practice at work in testing theory so that we can, as with all strong adaptations, know where we can use the common practice and where we must push and challenge it. After restating the specific focus and research questions of this study, the current chapter goes on, first, to discuss some of the literature on validation to outline what is involved in such work inside the argument framework of validity. In particular, I outline the general structure of a validity argument by primarily discussing Kane's (2006) current explication of validity in the fourth edition of *Educational Measurement* and the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) *Standards for Educational and Psychological Testing* (1999),¹ showing that the language and practice of validation reveals that we must design studies based on our particular contexts.

Second, I detail the particular context of Cardinal College as well as the *interpretive argument* (Kane, 2006, p. 22) for our writing placement process as I discuss the specifics of such aspects as my background, the College population, our courses, our teachers, and our placement process, which includes our placement prompt and reading

¹ The 1999 standards were the most current standards during planning and primary data collection for this study, which took place from August 2013 until January 2014. An updated *Standards for Educational and Psychological Testing* was published in early 2014.

method. Third, I describe a key assumptions inherent in our interpretive argument that this study addresses in its research question. I outline both why this assumption is the most pressing for study and how future studies will have to continue the work of validating Cardinal College's program by examining other pieces of the interpretive argument. Fourth, I outline the specific validation requirements and evidence that were required in an investigation of the assumption of the interpretive argument that is the focus of this study. As I do so, I make an argument for the primarily qualitative nature of this evidence given the particular context of our program, its assumptions, and the importance of teaching and learning based assessment for the field, and I provide a rationale for the specific data collected and the analysis decisions I have made.

Fifth, I specify the data collection steps for the study with three phases of data collection and various evidence sources that include think-aloud protocols of placement readers and course teachers, interviews of the same, a survey of teachers' perceptions of placement adequacy, course documents, and copies of student work. Sixth, I detail the procedures for data analysis, including an outline of steps to organize and reorganize the sets of data according to relevant placement or course designations and to develop coding schemes or coding categories for the various data sets. For applicable data sets, I include a description of the resulting coding schemes or categories developed. Then, I go on to detail the steps taken to compare the data within and across data sets. Seventh, I discuss the strengths and limitations of the data collection and analysis of this study before concluding the chapter.

Focus of the Study and Research Question

Participating, as this study does, within the aforementioned newer definition of validity (Cronbach, 1988; Kane, 2006; Messick, 1989a, 1989b), the focus of the study and its research question represented beginning work toward the validation of one particular college's placement process. As outlined in Chapter One, the focus of this research overall was on studying the validity of the placement process at Cardinal College—a small, private college in the suburbs of a mid-sized, Mid-Atlantic city—which uses Smith's (1992/2009, 1993) *expert reader model* to place students in reading and writing integrated courses. In Smith's model, an expert rater is one who has recently taught the course for which they are making placement decisions, and in his findings, these raters more reliably and adequately make placement decisions by employing their teaching and learning expertise to look not only at students' writing, but at the undertaking of teaching these students.

Given this focus, this study answered the following research question:

- Do expert readers at Cardinal College use curriculum/course related and teaching specific criteria to offer a pedagogically useful and purposeful, view of students' placement essays? Are expert readers' criteria (and by extension their decisions) linked to course criteria?

Specifically, answering the question involved considering four subquestions essential to a consideration of the main research question:

1. What are expert-readers' placement criteria?
2. Are expert-readers' placement criteria teaching specific?
3. Are expert-readers' placement criteria linked to course criteria?

4. Are expert-readers' placement concerns and decisions linked to course concerns and decisions?

Pursuing the research question and its subquestions, this study began the process of crafting a *validity argument* for Cardinal College's placement by specifying the program's *interpretive argument* (Kane, 2006) and investigating a key assumption of the writing program's argument—that placement readers place students according to course content criteria.

Overview of Validation

Since the purpose of this study was to develop a validation argument for Cardinal College, it was important to first understand generally what is involved in validation. In what follows, I outline two parts of validation—the interpretive argument and the validity argument—and describe each. Next, I explain the contextual nature of validation work by discussing how the particular interpretive argument dictates what is required in validation studies.

The Interpretive Argument and the Validity Argument

According to the AERA, APA, and NCME (1999) standards, “Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). This definition makes clear that to validate involves two different pieces of work: outlining the proposed interpretation and use of a test and evaluating the proposed interpretation. These two pieces of validation were also evident in Kane's (2006) discussion of validity in the fourth edition of *Educational Measurement*. Kane (2006) wrote, “Validation always involves the specification (the interpretive argument) and evaluation (the validity argument) of the proposed interpretation and uses

of the scores” (p. 22). Here, Kane and the standards draw on the history of the current view of validity as an argument that can trace its roots to construct validity (as I have discussed in Chapter Two). When the view of separate distinct kinds of validity (predictive, concurrent, content, and construct) gave way to a unified view, many of the guidelines for construct validity specifically became principles in the newer view of validity as an argument. Specifically, Cronbach & Meehl (1955) were the first to suggest that the proposed interpretation of a test must be detailed before it can be evaluated, and their expression of this principle in construct validity survives in the unified view of validity as an argument that exists now.

Outlining more specifically what is involved in each part of a validation argument, Kane (2006) wrote, “An *interpretive argument* specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances” (p. 23, emphasis in original). He continued, “The *validity argument* provides an evaluation of the interpretive argument” that checks that “the interpretive argument is coherent, that its inferences are reasonable, and that its assumptions are plausible” (Kane, 2006, p. 23, emphasis in original).

Contextual Nature of Validation

While validation is described as this two-part process of specifying an interpretive argument and of evaluating it to create the validity argument, the current literature on validation provides little else in the way of specific guidelines. The different contexts and particular uses outlined in a particular interpretive argument have far more to do with what is necessary of validation at a specific site than general rules or principles. The

development and specification of the interpretive argument is actually what determines the interpretations and assumptions that need to be evaluated to craft the validity argument. Both Kane (2006) and the AERA, APA, and NCME (1999) standards discussed the way that an interpretive argument sets up what is needed in a validity argument. In the AERA, APA, NCME (1999) standards, the particular context of a test and its use—specified in the interpretive argument—reveal the assumptions and interpretations that a test rests on, and these assumptions and interpretations provide a “conceptual framework” for considering other possible explanations, deciding what to study, and determining what evidence is necessary to investigate the proposed interpretations (p. 9).

Because the work involved in validation at a particular site emerges from that site’s particular interpretive argument, the validation process and the evidence required will differ across sites and contexts. By articulating the goals of a test, the assumptions behind its creation, and the inferences and interpretations that are involved in using a test, we uncover what will be needed to study a particular program. For one particular site, a certain kind of evidence may be essential to support the proposed use of a test, but at another site, that same type of evidence may be unnecessary (AERA, APA, & NCME, 1999).

Since the work of validation and the specific evidence necessary for it are contingent on the context of a particular case and the interpretive argument behind the use of a test in that case, a lot of responsibility falls on those using a test to conduct validation research about that test’s use by outlining assumptions, by being continually skeptical or critical of intended interpretations, by considering alternative interpretations,

and by studying the consequences of a test specifically. To develop and specify the interpretive argument and to evaluate it during the creation of the validity argument naturally involves an ongoing program of study (Cronbach, 1971) since it is not possible to investigate all the assumptions behind the interpretation of one test's use in one study. Additionally, it is not possible to effectively tease out all possible alternative interpretations in one study, which Kane (2006) stated is part of validation—an idea originally implied in Cronbach and Meehl (1955) and Cronbach (1971), but explicitly stated by Messick (1989b). Nor could one do all this—specify the interpretive argument, evaluate it, and investigate alternative interpretations—in a single study while also examining the consequences of a test's use (Cronbach, 1971, 1988; Messick, 1975, 1989b, 1995), another requirement of validation. Further, since Kane (2006) differentiated between the validation work done during a “development stage” and that done during an “appraisal stage” (p. 25), an ongoing program of validation is preferred since early studies, done when a test is in development, can focus on specifying the proposed interpretation without needing to be as critical as later studies which can turn to the critical evaluation of stated interpretations.

The explanation above, regarding validation and how its work is linked to the particular context under study, was influential to the design of the study of Cardinal College's placement process and the contents of this chapter. To understand what would be necessary in validation, we first had to understand the argument we make for our assessment practices. Therefore, this chapter goes on to specify Cardinal College's context and the arguments inherent in its placement process—its interpretive argument—and it uses the resulting description, along with additional writing assessment literature,

to provide a rationale for which piece of Cardinal College's interpretive argument was selected for study, as well as for the data collected and analyzed in the study.

Outlining Cardinal College's Context and Interpretive Argument

Now that we had a sense of validation generally, it is important that we understand the particular context of Cardinal College. Additionally, Kane (2006) made clear that a validation argument is built only after a specific interpretive argument has been specified because the process and act of validation is specific to the particular argument being made for the use, interpretation, and decisions of an assessment. Therefore, in the following section, I detail specifics of Cardinal College's context, including my own background as an adjunct teacher in the writing program, as well as details of the student population, courses, teachers, and placement process. As I offer some sense of our context, I am putting forth the interpretive argument for our courses and their connection to our placement program.

My Background

I approached this research study not as an outsider; I am a member of the community. I have a close relationship with Cardinal College and the writing program at the College. It is important that I outline my own experiences with the program so that it is clear that I have intimate knowledge of the context and have developed my own interpretations and views that cannot help but impact what I see, and therefore should be made public and explicit (Creswell, 2007). Since the fall of 2004, I have been an adjunct faculty member working in the writing program. I have taught all courses in the sequence at various times and for multiple semesters during the nearly 12 years I have worked as an adjunct instructor. I have also taught a few other courses outside the freshmen writing

program on such topics as linguistics, reading and writing across the disciplines, and business writing. Additionally though, I was a student at the College directly out of high school, and as I worked toward a bachelor's degree in Secondary English Education, I took the required writing program courses at the College, and eventually worked to help other students meet these requirements as I became a consultant in the College's writing center, an ESL tutor, and a writing program teaching assistant.

The Student Population

Since Cardinal College and its particular first-year students taking the writing placement exam made up the population for this study, it was important to detail more about that particular population. Cardinal College is a small, private Catholic college in the suburbs of a city in a Middle-Atlantic state. Accredited by the Commission on Higher Education of the Middle States Association of Colleges and Schools, the college awards associate's, bachelors, and a few master's degrees. Numbers from a report on commonly requested data from the office for institutional research for Fall 2013 show important information on the population of the college useful in understanding its particular context and makeup.

For instance, the College primarily serves undergraduates since the total enrollment for Fall 2013 was 1,481 with 1,363 of those being undergraduate students. Further, during the semester of study, the undergraduate population was 57% female and 43% male, with nearly a 13% international student population. The international population was made up of students from 33 countries with the largest numbers of students from such countries as Saudi Arabia, China, Korea, Canada, Burundi, and Kenya. In addition to the nearly 13% international student population, 82% of the total

undergraduate population came from the same state and 5% from out of state, drawing from 16 states and 2 territories with the largest numbers of students primarily from contiguous states or those on the Mid-Atlantic seaboard. Of the 1,123 students from the same state, 1,010 of them were from the same corner of the state as the College, with 728 of those from the same county as the College. There was a small group of transfer students as well with 167 students enrolled at the College in Fall 2013 who transferred credits from another institution. Over the last few years, these numbers have remained fairly consistent although there are always slight variations. For example, comparing these numbers to Fall 2011 shows that the Fall 2013 number of international students was up 3% while the number of transfer students was down nearly 3%.

Since this study dealt with the writing placement test directly which is taken by first-time, full time freshmen, the breakdown of these students is important to distinguish them from the larger group of undergraduate students. During Fall 2013, the first-time, full time population was also 57% female and 43% male, but with a larger international student population at 17%. The first-time, full time international population was made up of students from 16 countries with the largest numbers of students from such countries as Saudi Arabia, China, Kenya, Korea, and Vietnam. In addition to the 17% first-time, full time international student population, slightly more than 72% of the total first-time, full time population came from the same state and slightly more than 10% from out of state, drawing from 9 states and 2 territories with the largest numbers of students coming primarily from contiguous states. Of the 166 first-time, full time students from the same state, 143 of them came from the same corner of the state as the College, with 96 of those from the same county as the College. While the average undergraduate student age during

Fall 2013 was 23, the average age of first-time, full-time freshmen was 19. All entering students who do not come with writing transfer credits must take the writing placement, and since the population includes some adults and the aforementioned international students (who are usually a little older than traditional first-time, full time freshmen), the average age is pushed up. However, while the average was 19, most students sitting for the placement over the summer were 17-18 years old.

First-time, full-time students applying for admission to the College submit high school GPAs and SAT scores meant to give the admissions office some indication of their academic readiness for college. In Fall 2013, 74% of entering freshmen submitted their high school GPA, and the average GPA was 3.186. Yet, 34% of these entering students had under a 3.0 high school GPA. Further, the average critical reading score for the 69% of students who submitted SAT scores that semester was 454. This average is 42 points under the national average of 496 (Snyder & Dillow, 2015). Comparing this to Fall 2011 figures shows our student population consistently falls under the national average: In Fall 2011, 80% of students submitted SAT scores, and they averaged a 465 SAT critical reading score, which was 32 points under the national average of 497 for that year (Snyder & Dillow, 2012). These numbers can be attributed to certain factors like the recruitment for and marketing of a program at the College to admit and provide extra mentoring and assistance to at-risk or underprepared college students, and many of Cardinal College's students are first-generation college students.

Looking at first-year students, those who take the writing placement test, can give a sense of how many students typically make up the population of writing placement takers in a given fall semester. During Fall 2013, there were 229 first-time, full-time

freshmen. 205 actual students sat for the writing placement during the summer prior to Fall 2013, and the total number of students enrolled in a first semester course (Course 10 or Course 11) during Fall 2013 was 242. The discrepancy in numbers here is due to a number of factors. First, while there is a small number of students who attend a placement testing session but who do not go on to actually attend the college, there are far larger numbers of students who move up to first-time, full time status from the English-as-a-Second-Language (ESL) program. This means the number of first-time, full time freshmen is usually larger than the number of students who sat for the placement exam during the summer sessions since ESL students take the placement exam inside their final ESL advanced writing class. In Fall 2013, 19 ESL students took their placement exam this way. Additionally, the number of students enrolled in a first semester course is larger than the total of first-time, full time freshmen because the enrolled number includes those who retake a course after failing it previously and those who have moved up to the higher incoming freshmen writing course from a prior semester taking the lower one.

Courses

Given the student population—with its large number of international, underprepared, and first-generation college students—the writing program at the College stresses the development of academic literacy skills in both reading and writing to give students an “opportunity to learn” in a discourse that is potentially quite unfamiliar to them (Gee, 2003, p. 1). Additionally, many students at the College arrive in our writing courses with a few distinct reading characteristics that have made the focus on reading and writing necessary. First, they are often quite unfamiliar with reading for purposes

other than summary or information gathering (Bartholomae & Petrosky, 2004). Second, they struggle to develop their own understandings of texts because they are reluctant to trust in their own interpretive authority (Bartholomae 1985/2005; Sommers & Saltz, 2004), preferring instead to be told the meaning or significance of texts. Third, they do not approach reading as if it is an act of composing like writing (Tierney & Pearson, 1983), with the possibility for seeing new information and coming to new understandings through subsequent work on the text and the text of their own reading (Bartholomae & Petrosky, 1986, 2004). And fourth, when approaching their own writing or the texts of fellow classmates, they do not approach these texts (by applying similar reading strategies, or reading as closely and critically) as they do published course texts (Bartholomae & Petrosky, 1986; Coles, 1988), a difficulty that leads them to struggle with revision (Flower, Hayes, Carey, Schriver, & Statman, 1986; Larson, 2000).

Drawing largely on the work of David Bartholomae (1979, 1985/2005, 1993) and David Bartholomae and Anthony Petrosky (1986, 2004), the writing program is crafted specifically to help Cardinal College's student population prepare for academic reading and writing generally, as well as recognize the variety of choices, strategies, approaches, and discourses inherent inside the academy. It is important to understand the conception of each course the College requires to get a sense of the argument behind how each is envisioned and how they fit together. To begin, I discuss the two, three-credit courses that all students (except those with transfer credits from another acceptable institution) are required to take: Course 11 and Course 12.

Course 11. Course 11 is designed to introduce beginning academic writers to the interpretive reading of complex academic texts, including the process of developing a

reading and a revision of that reading through their writing (Bartholomae and Petrosky, 2004). Further, the course seeks to help students begin to take part in the kinds of conversations about texts that are taking place in the academy (Graff, Birkenstein, & Durst, 2012). Assignments in the course have students learn to take a position in relation to other texts—a position in which they have a stake, and a position that they must explain and support. In the process, they learn to put texts into conversation with each other—including the text of their own experience (Sommers, 1993)—to more fully understand these texts and their own emerging positions.

Course 12. Course 12 extends the work of Course 11 by introducing students to academic research. The course requires students to investigate academic texts and conduct research with an eye toward understanding the choices and decisions of writers (Bunn, 2011) and to consider what research means academically, including an acknowledgement of differences between popular and scholarly research, between primary and secondary research, and across disciplines. Students carry out this investigation by examining closely how other writers conduct research and make meaning using their materials or sources and by completing several essays on their findings. The process of learning about research is finalized when students research, draft, revise, edit, and present their own significant research project using what they have learned from their examination and discussion of other researchers.

Course 10. All students at the College must complete Course 11 and Course 12, but since students entering the composition sequence have different abilities, and some students may not be ready for Course 11, the program has an additional placement choice: Course 10. Course 10, a six credit class (where all credits count toward

graduation and financial aid), is designed to prepare students for the kind of reading and writing found in Course 11 by immersing them in intensive reading and writing activities (Bartholomae & Petrosky, 1986). Rather than approach students in the traditional way of developmental courses where students are seen as requiring remedial work on such things as grammar skills or writing effective sentences (rather than whole compositions) (Otte & Mlynarczyk, 2010), students are given assignments nearly identical to those in Course 11, but they are given more instruction, scaffolding, and modeling throughout the reading and writing processes on such things as reading strategies, note-taking during reading, developing a position in relation to a text, essay structure, the presentation of quotations, practicing peer feedback and self-assessment, and identifying and controlling error.

Since Course 10 has the same basic goals as Course 11 but has twice as many contact hours/credits, it is possible that students placed in Course 10 who do very well can potentially skip Course 11 and go to Course 12 with the recommendation of their instructor. Yet, since Course 10 is also a way to assist students who may need significant development and support, if upon completion, students need more practice and development, they could continue onto Course 11. The number of Course 10 students moving to Course 11 versus the number moving to Course 12 varies slightly across semesters, as the student population changes, but to get a sense of the numbers, during Fall 2011, the director of the writing program reports that 68% of the students who placed into Course 10 went on to Course 11, 17% of the students moved directly to Course 12, and 15% of the students failed Course 10. In Fall 2013, the semester of study here, the numbers were similar. 76.06% of the students placed into Course 10 went on to Course

11, 13.68% of the students moved directly to Course 12, 8.55% of the students failed Course 10, and 1.71% of the students withdrew from the course before finishing.

Course 11A and Course 11L. At one point, the College had one additional placement option: Course 11A. This course was identical to Course 11 in objectives and content, and counted identically in completion of the three credit, Course 11 requirement; the difference was that the course was taught with an additional hour per week (bringing the total to 4 hours per week) for specialized time spent on developing close reading skills and working on error control for students who were deemed in need of some additional scaffolding but not a full six credit course. In Fall 2008, due to space limitations that required courses offered at the College to be a consistent three hours per week or one hour per week for ease of scheduling, the Course 11A option was eliminated (a decision made by the administration).

In the place of Course 11A, the writing program negotiated for all students except for transfers and those in special programs (which could not accommodate an additional credit) to be required to take a one-hour lab course in addition to their Course 11 course (scheduled separately, and with a separate teacher). Course 11L, which was pitched as stressing reading and writing in a timed environment, does help students learn strategies that will help them complete reading and writing tasks more quickly. However, its reiteration of Course 11 concepts—where students learn to read to locate “academic conversations” (Graff, Birkenstein, & Durst, 2012, p. ix), and where students participate in an active discussion of the reimagining, revision, and editing of previous work—gives all students—in particular the old Course 11A population—additional practice with Course 11 concepts within the new framework of on-demand writing.

Writing Program Teachers

Teachers in the writing program are pulled from the English department, as well as from a pool of recurring adjunct instructors. There are seven full-time English faculty members at the College who must teach four courses per semester in addition to advising duties. Occasionally, during a particular semester for instance, a full-time faculty member may not teach in the writing program if their course load requirements are met by other required courses. In fact, there is one member of the full-time faculty who very rarely teaches in the freshmen writing program since his expertise and duties running the College newspaper and teaching journalism, technical writing, and creative non-fiction writing assign him elsewhere. The rest of the full-time faculty split their course loads between courses in the freshmen writing program and other specialties. Specifically, some faculty teach required core courses in humanities or within one of the three distinct English majors in English education, professional writing, or language/literature.

Three of the full-time faculty members have English degrees with specializations in composition and rhetoric, writing, or literacy, while four others have backgrounds in literature, yet all have terminal degrees in their fields (6 PhDs and 1 MFA). The Chair of the English department, for example, has a degree in literature, but the Director of the writing program focused on writing and literacy. In addition to full-time faculty, there are nine adjunct instructors who are currently teaching in the English department at the College primarily in the writing program with a few also teaching online literature courses or specialty writing courses (like grant writing, for instance). They typically teach no more than two courses per semester in line with administrative policy. Like the full-time faculty, their backgrounds are mixed with two possessing degrees in English that

specialized in literature, three in literature and writing, one in English education, and one in reading. One adjunct professor has a degree in cognitive psychology with a focus on writing, and another has dual degrees—one in literature and the other in history. Four adjunct professors currently have a terminal degree, but three others, including myself, are pursuing a terminal degree or have pursued a terminal degree, reaching ABD status but, as of yet, not finishing.

Placement at the Site

Having outlined the particular context of our students, teachers, and courses—including the argument for the way our courses are envisioned, it is necessary to detail the argument for the placement process at the College. Currently, we ask students accepted to the College to attend a placement testing session where, in addition to a math and computer assessment, they also write an impromptu essay for writing placement. Below, I discuss specifics of the argument for our placement process, including a discussion of our rationale for using one direct, impromptu essay; an explanation and argument for our particular writing prompt; and the reasoning behind using an expert-reader model.

Direct assessment of an impromptu essay. Given the very real and clear debate on direct versus indirect writing assessment measures that has long been a part of the history of writing assessment (as I outlined in Chapter Two), it should come as no surprise that Cardinal College has not chosen to go with an indirect measure of writing ability. While the admissions office records students' SAT scores, and while the office in charge of placement testing has at times administered a grammar correction or reading comprehension test for its own records, placement decisions at the College are not made

using this information. While testing systems like ACCUPLACER and COMPASS now offer a computer-scored writing component in addition to their more traditional grammar or comprehension tests, Cardinal College has students, who are on campus for other placement tests in math and computers, write an impromptu essay on a locally written prompt that is later read and placed by teachers in the writing program.

There is real criticism in the literature on impromptu essays that the writing program at Cardinal College has to acknowledge and account for in its interpretive argument. For instance, Breland, Camp, Jones, Morris, and Rock (1987) discussed the need for more than one sample, written at more than one time, and in different modes to make decisions on writing ability, and this was an aspect acknowledged by the Conference on College Composition and Communication's (CCCC) (2009) position statement on writing assessment. And White (1995a, 1995b) and Purves (1995), in a back and forth conversation on the impromptu essay, both acknowledged serious issues with its ability to represent the whole of the writing process as it is taught in composition classrooms. Finally, Leki (1991) discussed ESL students specifically and mentioned, among other constraints, the influence of the time allotted for reading the exam and writing the response on ESL students' performance.

The discussion of these scholars outlines many concerns about the limits of impromptu essays, but some scholars also offer support to Cardinal College's own position. As White (1995a, 1995b) defended the impromptu essay and rebuffed Purves' (1995) response to that defense, he claimed and reiterated that while portfolio assessment is preferred whenever contextually and practically possible, an impromptu essay can be quite useful, if designed for the particular context, when time and money constraints

mean that portfolio assessment is not possible. White's practical treatment of decision making in regards to impromptu essays was backed up by Haswell and Wyche-Smith (1994) who also adapted Smith's (1992/2009, 1993) expert model to their context and argued, as White did, that given their "pragmatic constraints" with students only on campus for a "two-day orientation," the students were "too rushed and distracted for arduous testing, and [the] brief turnaround time for reporting results forbade complex samples" (Haswell & Wyche-Smith, 1994, p. 230). For this reason, Haswell and Wyche-Smith chose the impromptu essay as the best possible choice given the circumstances.

These scholars offer some support to Cardinal College's own situation and position. For the purpose of placement, it is difficult to ensure students have enough previous work to assemble a portfolio, and given their limited time and availability to be on campus, more than one writing sample is a practical nightmare. By combining the one sample at placement with a retest on the first day of class (essentially a second sample that rechecks placement), the writing program feels comfortable that the constraints of our context had been adequately balanced. Additionally, given the goals of our program, we have some confidence in making a decision based on a piece of writing that we argue is matched to our particular curriculum. However, even with these decisions, we still have had very real concern that for many students—ESL students in particular—insufficient time could seriously limit their ability to do well. Therefore, we have always allowed students to take all the time they felt they need to finish their responses (no student has ever exceeded two hours of writing), and starting this year, we have begun providing students with copies of the two short readings and the prompt in advance so they can read and think at their own pace prior to the placement session if they prefer.

Placement prompt. In 2007, our current prompt (see Appendix A) was created when it was felt by most teachers that the previous prompt, which asked students to read a short text on cheating in college and write their own response to the position of the author, was not revealing enough information about how students handled reading and writing tasks more complicated than summary and response. Many students could offer an acceptable response to one text, even those students with little academic reading and writing experience and who would go on to struggle with other course tasks. Thus, it was felt that the prompt did not solicit student writing that was helpful in determining how the students would handle critical, text-on-text work.

Since many Course 11 teachers frequently ask students in their courses to apply their understanding of one text to a scenario, example, or situation outlined by another (what teachers in the program characterized as a *lens* reading), the prompt was changed to reflect this work. We are supported in our decision by such findings as Smith et al. (1985) who concluded that different kinds of prompts worked better for distinguishing between different levels of students. Smith et al. gave students at various levels in their writing program prompts with three different structures—one in an open structure where students could discuss a topic using their own experience, one where they were asked to respond to one text, and one where students were asked to respond to several texts on the same topic. In the end, the differences between students at various levels were more noticeable for the prompt that included multiple texts. Smith et al.'s findings support our belief that adding an additional text to the task would better help us differentiate between basic writing and college writing students.

More than just the inclusion of readings though, the choice of writing task is a potentially problematic issue since, as Murphy and Ruth (1993) made clear, writing tasks can be and are interpreted in a number of ways by different students and different raters that may be quite different from the intended meaning of the test maker. This possibility for various interpretations is accented by Leki (1991) who discussed ESL students specifically and by Murphy and Yancey (2008) who talked about potential differences in interpretation for all students with linguistic or cultural backgrounds different from test makers. The claim, by Murphy and Yancey in particular, was that such variation of interpretation introduces *construct irrelevant* variance (Messick, 1989a) into the writing task. The assumption here was that how students interpret what the prompt is asking of them is not (and perhaps should not be) relevant to an assessment of their writing. However, such an argument only works if reading and writing are conceived of as separate in the curriculum.

Given Cardinal College's integration of academic reading and writing in the curriculum of the program, the students' interpretation of the prompt is not seen as irrelevant at all. It is the argument of the program that how students understand the writing task they are being asked to complete is potentially revealing about their familiarity with, and even the depth of interaction with academic reading and writing. For, as Gee (2003) argued, not all students have had the same *opportunity to learn* the kinds of reading and writing expected in specific social contexts, like the academic context of school. Further, the program feels that since the placement is read by expert teacher-readers, differentiations can be made between the kinds of interpretations of the

writing task that might interfere with a student's performance in Course 11 and those that will not.

Placement reading method. As discussed earlier, Cardinal College makes placement decisions following a modified version of Smith's (1992/2009, 1993) expert reader model. Smith's model developed from his reliability and adequacy studies into the University of Pittsburgh's placement process, and he found that teachers who shared recent teaching experience in particular courses would make more reliable and adequate decisions about placement into their course of expertise when they were making only a decision that the student belonged or did not belong in their particular course of expertise. Kane's (2006) description of teachers' *conceptual frameworks* used in classroom assessment also lends significant support to the idea that teachers' expertise was useful. He wrote,

Teachers bring a potentially rich set of tools to their observations of student performances. Experienced teachers meeting their new class for the first time have a good sense of the range of skill levels to expect. They know some of the gaps in understanding and skill to expect, and common impediments to learning, and are familiar with "the relevant experience and discourse patterns" (Shepard, 2001, p. 1075) of their students. Their conceptual frameworks provide templates for organizing observations of student performances and for differentiating critical issues from more routine observations. Using these frameworks, teachers interpret student performances as they occur and do not simply keep a record of their observations for later interpretation. (Kane, 2006, p. 47)

Teachers, as Kane described them, have relevant knowledge and experience which help them as they observe students work and behavior.

Essentially, the program has always followed Smith's (1992/2009, 1993) expert reader premise but has used fewer readers. Prior to this year, the program had one reader who had recently taught Course 10 (and who had taught it more than once) read all essays and either place students into the course, or push them up. We had historically used one rater because Smith (1993) discusses how, given their agreement, two expert-raters can turn out to be redundant (p. 199), and since the program did not need to differentiate among students across four placements, as Smith did, it seemed practical to let one expert reader make a decision of rejection or acceptance into a course. The Course 10 rater was chosen because tapping that particular expertise seemed important for decision-making.

Over the last few years, due to a concern with student population changes and with the elimination of the Course 11A placement option, the Director and program have become interested in tracking potential *tweeners*, as Smith (1992/2009) calls them. Essentially, these are students who would be rejected high from Course 10 but would go on to also be rejected low from Course 11. Additionally, taking in Haswell and Wyche-Smith (1994), Harrington (1998), and Huot's (2007) procedure of placing the most populated, general course first since it presents the least difficulties to any reader, the procedure has been altered. An additional reader who has most recently taught Course 11 was added this year. After some experimentation during early summer placement sessions, it was decided that the Course 11 reader would read first and either place students in Course 11 or reject them low to be read by the Course 10 reader. This decision would allow the program to track the number of tweeners, to collect data on

their performance, to make an argument for future course elimination/creation decisions, as well as to discuss how best to place these students into the existing program. And, it would mean that two readers would have read each Course 10 placement, which is ideal given the consequences associated with inadequate placement into a six-credit course prior to starting the required college writing course.

Key Assumptions of the Interpretive Argument: Selecting the Study Focus

Given our specific context and the specific requirements for constructing a validation argument based on our context, there are a number of assumptions inherent in our interpretive argument, and a number of other factors such as the consequences of assessment decisions that would need to be investigated in any validation. As I discussed earlier when I outlined the literature on validation, it is outside the scope of one study to investigate all assumptions of an interpretive argument as research into validation should be ongoing and continual (Cronbach, 1971). Going forward, I detail two key assumptions uncovered in the interpretive argument, and explain which one this study investigated, making a case for why it was critical for our program to study.

The first assumption is that expert teacher-readers read placement essays using curriculum and course related criteria, as well as criteria specific to the concerns of teaching students in the context of specific courses. Operating, as we do, from Smith's (1992/2009, 1993) expert reader model, we have never investigated, as Smith did to arrive at his model, that placement readers were using such criteria. If we want to assert, as Smith did, that our placement reading has value and usefulness for making adequate decisions about students' course placements and for giving us a window into the teaching of our students, we have to back up this claim with data. Since the expert model is a

lynchpin piece of our entire placement process, the investigation of this assumption is key. It is also an important assumption to investigate given recent and continual calls from the administration to consider using a computer-scored, online system like ACCUPLACER or COMPASS in place of our locally-designed system.

The second key assumption of our placement process is that the placement prompt is intimately related to the construct of reading and writing represented by our courses, and therefore assists our expert placement readers in making a course-related decision about placement. As I detailed above, the prompt was changed when it was felt by many teachers that the previous prompt was not allowing placement readers to see which students may struggle with more complicated reading and writing tasks. The choice of task, however, has never been investigated empirically. Does the prompt represent consistently the work of the courses, or only some teacher's conceptions? Does the task of a lens reading expose difficulties that would affect students' performance in Course 11? Are related but preceding abilities developed in Course 10 that may not be sufficiently addressed in Course 11? These kinds of questions can only be answered with study. Since it is possible that a poorly chosen task can place students into the lower course unnecessarily, a concern raised by some administrators and the staff in the placement testing office who feel that the task of the prompt is too difficult, investigating this assumption is important to validating Cardinal College's interpretive argument as well.

While both assumptions are important for Cardinal College to study, it was outside the scope of a single study to fully consider both aspects. Given that the findings from investigating the first assumption—that readers are using course-related criteria to

determine placement—will naturally provide a sense of how useful the prompt is for placement readers, this study focused on the first assumption, using insight gained from research to plan future study on the prompt more directly.

Rationale for Study Design

Given the main assumption that stems from the interpretive argument of Cardinal College's writing placement process—that placement readers use course criteria at placement—it was clear that to begin crafting the validation argument, this study had to gather data to investigate this assumption. What follows details the logic behind the design plan of this study. As I go, I offer a kind of miniature review of validation literature and relevant literature on particular research methods. My goal in offering a review of these sources is to provide a rationale for how the study was designed to evaluate Cardinal College's interpretive argument.

First, I describe two specific validation scenarios discussed by Kane (2006)—placement testing and classroom assessment. Using the descriptions offered by Kane, I develop a rationale for the decision to focus on classroom assessment over the standardized placement analysis, as well as a rationale for conducting a primarily qualitative study. Next, I review specifically the data that Smith (1992/2009, 1993) discusses as useful for determining adequacy of placement. Since, as I discussed earlier, Smith's expert-reader model is used by Cardinal College, I emphasize Smith's data and Kane's classroom assessment description over Kane's description of validating placement. In the review of Smith's adequacy data sources, I offer a rationale for the choices of adequacy data and the specific forms the data in this study take, specifically think-aloud protocols, teacher interviews, and course documents.

After detailing the choice of data sources for the study, I discuss the general theoretical approach I would take to coding data, as well as coding models that I was familiar with from the literature which could influence my thinking. Then, I review the AERA, APA, and NCME standards in regards to kinds of evidence for the work of evaluating an interpretive argument so that I can show that despite an emphasis on qualitative data and classroom assessment, the data plan still meets the guidelines offered in the standards. Finally, I end with a summary of the approach, data, and analysis plan sketched in this section.

Evidence for Placement Testing and Classroom Assessment From Kane

Because each interpretive argument is specific to the particular context and use of a test, the validity argument that evaluates it is also specific to that context and use. Yet, while the specifics in each case will differ, Kane (2006) illustrated how to outline and evaluate an interpretive argument by discussing a few common school assessment undertakings. Specifically, he described the inferences and assumptions common to placement testing, as well as those related to the context of classroom assessment. While his purpose was partly to demonstrate the task of specifying an interpretive argument in scenarios that many readers would find familiar and instructive, he was careful to show that not every instance of, for example, placement testing is identical. The guidelines Kane (2006) gave in relation to placement testing were important to consider since it was exactly the situation of placement that was the focus of this study and its research question.

Placement testing. To begin his description of an interpretive argument for placement testing, Kane (2006) specified the goal of placement testing which he said,

is to assign students to courses that will be optimal for them in some sense (i.e., to a course that will be demanding but not overwhelming for the student). The ideal course would be one focusing on competencies that the student has not mastered but for which the student has mastered all prerequisites. (p. 24)

Kane's description of the goal revealed key assumptions specific to placement testing that seem to be shared across instances of specific placement testing. This description suggested that those administering placement tests assume that their testing procedure (the test, scoring, etc.) offers a way of differentiating between students who would do best in different available courses, that the courses develop the skills necessary for mastery of a subject, and that there is a developmental relationship between available courses.

The goals specific to the task of placement led Kane (2006) to outline four specific inferences he saw in placement testing, the details of which will look quite different across different particular contexts even as the basic assumptions are the same. He first discussed the *scoring inference* of placement testing which details and describes how a placement program will move “from observed performance to an observed score” (Kane, 2006, p. 24). Part of the scoring inference is the assumptions that the way tests are scored is “appropriate” and “applied accurately and consistently” (Kane, 2006, p. 24). As he discussed each inference, Kane turned to describe some of the ways such inferences are evaluated. A scoring inference is evaluated through studies which typically involve an expert's judgment of the appropriateness of the scoring rule or a statistical determination of the agreement of raters (Kane, 2006, p. 25). The particulars of the scoring rule in each

system of placement will, of course, vary, but the key for Kane was that an important inference of placement is the movement from a performance to a score.

Kane (2006) next discussed the *generalization inference* which he said helps move “from observed score to universe score” (p. 24). In short, the generalization inference deals with how we know the performance on this particular measure is representative of the larger domain of performances. Kane (2006) said that such an inference assumes that “the observations made in testing are representative of the universe of observations” and that “the sample of observations is large enough to control sampling error” (p. 24). The testing of this assumption is done primarily through reliability and generalizability studies, although expert judgment that “the sample of observations included in the test” are representative can be secured as well (Kane, 2006, p. 25). Reliability studies seek to readminister the measurement to ensure that performances (or scoring decisions, in the case of rater reliability) are repeatable; generalizability studies seek to understand how well a particular measurement represents the universe of measurement by looking at different factors that can affect performance, such as room temperature, time of day, or order of test items (Kane, 2006, p. 35).

The third assumption Kane (2006) discussed was the *extrapolation inference* which moves “from universe score to the level of skill” and assumes “the test tasks require the competencies developed in the courses and required in subsequent courses” (p. 24). In addition to assuming the test relates to the skills being assessed, a part of this inferences involves an assurance that “skill irrelevant” factors are not impacting scores casting doubt on the measurement as an assessment of the skill under study (Kane, 2006, p. 24). Kane (2006) said the evidence gathered to evaluate the extrapolation inference

usually involves looking at “the overlap between the skills measured by the test and those needed in the courses” either by securing expert judgments or conducting “empirical analyses” that connect the test score to “course grades” (p. 25).

The final inference of placement testing that Kane (2006) discussed was the *decision inference* which, as he described, moves from our “conclusion about level of skill to placement in a specific course” (p. 24). To make such a move, those administering placement tests assume, and therefore must show, that “performance in courses, beyond the initial course, depends on level of skill in the competencies developed in earlier courses in the sequence” (Kane, 2006, p. 24). To claim such a developmental relationship between courses means those administering the placement test believe that that students who have not shown mastery of “prerequisite” skills for a certain course “are not likely to succeed in the course,” and that students who have shown mastery of the skills taught in a certain course “would not benefit much from taking the course” (Kane, 2006, p. 24). Since it is clearly unethical to purposely misplace students in order to assess struggle or success, evaluating the decision inference involves looking at “the positive and negative consequences resulting from the decisions” and how those consequences relate to the consequences of other possible decisions (Kane, 2006, p. 25).

Classroom assessments. While Kane’s (2006) outline of the inferences involved in placement may appear to be all that is necessary to begin detailing the specifics of the research plan and data in this study, there was one additional aspect Kane discussed that is important to outline, and that aspect was the nature of teaching and classroom assessments. A large part of our writing program’s placement process rests on an assumption that expert teachers’ course knowledge and their skills at classroom

assessment allow them to make adequate placement decisions by helping them recognize student writing and behavior typical of early course performances in their course of expertise. Given this key assumption, some knowledge of the work of classroom assessment seems pertinent to our context.

Using an example of a teacher watching a second-grader “read a story aloud” while “stumbling over some more difficult words,” Kane (2006) illustrated a teacher’s ability to make sense of a lot of data inside the specific theories, understandings, materials, and goals that make up teaching in a particular context (p. 47). Since the second-grade teacher Kane described comes to believe that the student is not unable to read, but instead that their “vocabulary is being stretched by the story,” Kane (2006) spent a good deal of time outlining the idea of a teacher’s “conceptual framework” as explanation of how teachers arrive at their interpretations of students’ performances (p. 47). To Kane, a conceptual framework involved knowledge, experience, and understandings that the teacher possesses that he or she uses to interpret students’ performances. In the case above, Kane showed that the teacher has knowledge of reading theory and development and, therefore, understands that unfamiliar words can slow reading, even for typically competent readers. Additionally, Kane (2006) discussed how the teacher’s familiarity with “the difficulty level of the text,” as well as with the student, can add weight to such an interpretation of performance—that is, if the text is known to have vocabulary students usually find unfamiliar, or if the student typically does not struggle when reading other texts, then the interpretation that the teacher has arrived at is strengthened (p. 47).

Through this description of teachers' expertise at work, Kane (2006) made two points about the work of classroom assessment that become important to understand in relation to Cardinal College's interpretive argument. The first point that emerged in Kane's description of classroom assessment was that such assessments are qualitative in nature. As Kane detailed the work of placement validation, he described how many of the inferences of placement were frequently evaluated through evidence based in the standardization and statistics of quantitative assessment. However, teachers' classroom assessments are not typically quantitative. Kane (2006) wrote,

Different kinds of information from different sources are combined for an interpretation of performance in context (e.g. that of a student in a class). Instead of the observation-scoring-interpretation paradigm prevalent in standardized testing, qualitative assessment involves an active search for meaning from the beginning, with the interpretation being elaborated and extended as data are collected....The qualitative approach focuses on evolving interpretations of observations rather than on scores. (p. 47)

Instead of standardizing their assessments by trying to control "variables" or generalize across them, teachers use qualitative assessments as they gather details, descriptions, and understandings and seek to make sense of how these data interact (Kane, 2006, p. 47). Citing Moss (1994), Kane (2006) asserted that a teacher's understanding is built as he or she attempts to bring all the information into a coherent whole and is continually revised as new data emerges (p. 47).

The second point that came out of Kane's (2006) discussion of the classroom assessment of teachers stemmed directly from the first. If teachers' classroom

assessments are qualitative in nature and do not rely on statistical standardization, how do we evaluate such assessments? We can no longer follow the movement from scoring to decision Kane has described for placement testing, a movement that, like other standardized measures, depends on the generalizability of the score which is important in statistical standardization. What then must be done? With his focus on the qualitative assessments of teachers, Kane (2006) said the same structure of validation still applies where an interpretive argument must be outlined and then evaluated; however, “the plausibility of the teacher’s conceptual frameworks and of the evolving views of students” (p. 48) is what is evaluated. To evaluate a teacher’s conceptual framework, Kane (2006) explained that the teacher must detail their interpretations by explaining what evidence they have for the understandings at which they have arrived, but the teacher’s “conceptions and organization of subject matter” as well as their “pedagogical theories and techniques” would also need to be detailed (p. 49). Finally, the teacher’s “training, credentials, and experience” should be given so there is additional evidence that the teacher has the qualifications to make such determinations (Kane, 2006, p. 49).

Reasoning Behind a Qualitative Approach

The decision to take a primarily qualitative approach was directly related to the interpretive argument of Cardinal College’s placement program and our own research goals. Our program’s approach to placement testing does not look like Kane’s (2006) description of the validation needs of placement testing generally—with scoring and generalizability inferences (p. 24) that require statistics and standardization in validation. In placement, Cardinal College does not produce a score at all, but instead, we make a placement decision directly. Additionally, our approach, with its focus on the decisions of

expert teacher-readers inside our local context, is more like Kane's (2006) description of the qualitative classroom assessment done by teachers where teachers use their "conceptual frameworks"—understandings developed from field specific knowledge, pedagogical theory, and past experiences with specific students—to make sense of current observations (pp. 47-49).

Since the focus was on our local context and our teachers' expertise inside that context, investigating our inferences and assumptions could not take a statistical or quantitative approach, but instead, required a qualitative approach which, as Huot (2002) indicated, should fit our focus on locally controlled assessments. He called for such an approach as he wrote:

We must...develop procedures with which to document and validate [local writing assessments]. These validation procedures must be sensitive to the local and contextual nature of the procedures themselves. While traditional writing assessment methods rely on statistical validation and standardization that are important to the beliefs and assumptions that fuel them, developing procedures will need to employ more qualitative and ethnographic validation procedures like interviews, observations, and thick descriptions to understand the role an assessment plays within a specific program or institution. (Huot, 2002, p. 106)

If we were to better understand how our assessments work in our local contexts, we needed something quite different than numbers. As Murphy and Ruth (1993) made clear in regards to the creation and study of writing tasks, statistics can indicate an issue or problem *is* happening, but qualitative methods can give us an understanding of *what is* happening (p. 277). In their case, using qualitative methods to watch teachers, students,

and test makers interpret writing tasks revealed different understandings of the writing task that were at odds with each other and influenced the reading of student's writing. In Cardinal College's case, while the planned methods would provide evidence that answered the research questions (say for instance, how the course criteria and the placement criteria were related), these qualitative methods would also give us something more. In short, we wanted details of the relationship between placement and the course that would help us name, make explicit, and refine our goals for teaching and learning.

Given the qualitative approach, Cardinal College would naturally be examined closely. Thus, it is important to mention that the goal of this study was not to use Cardinal College to find universal issues and approaches applicable to all. Since writing assessment scholarship acknowledges that there are no universal assessment answers to be found that will work for all, as teaching and learning are context-driven activities (Huot, 2002), the qualitative research methods of this study—informed by constructivist theories about knowledge that take as their beginning assumption that reality is not objective, but subjective and multiple (Creswell, 2007)—fit well with theories of assessment that informed the study design. Even though the goal was not to find universal truth, the qualitative description and discussion of Cardinal College would be used to aid the program in checking the effectiveness of its own decisions and in understanding more fully its own work.

Evidence for Determining Adequacy From Smith

Given the focus at Cardinal College on placing students following Smith's (1992/2009, 1993) expert reader model, some review of his discussion of determining the adequacy of placement is in order. Reviewing Smith is additionally helpful since Smith's

theories on teachers' expertise line up well with Kane's (2006) description of classroom assessment and teachers' conceptual frameworks.

Smith (1993) discussed five sources of data that are useful for determining the "adequacy" of placement, and they were:

1. The number of students who are moved to a new course at the beginning of the term because they were misplaced;
2. The final course grades of the students in each course;
3. The students' impressions (during and after the course) of the degree to which the course met their needs;
4. The teachers' impressions of how well the students fit into their courses; and
5. Exit exams (or other "posttests"). (p. 156)

Smith went on to describe each and his description made clear the logic of how these sources relate to the adequacy of placement. A program could investigate how many students were moved at the start of a given semester to have a count of who was misplaced, and it could look at final course grades to see how students did in the course. Additionally, exit tests could provide another measure for comparison purposes. But additionally, asking teachers or students themselves about their impressions on the fit of the course to the student could be an additional check.

Reasoning Behind Specific Data Sources and Study Plans

After naming and describing these sources, Smith discussed the values and problems with each kind of data individually. He also stressed the importance of "triangulating" the data to eliminate issues with individual data sources and reveal more information than a single source alone (Smith, 1993, p. 156). Going forward, I want to

review what Smith said about each potential data source, putting them in the order of their importance to this study, to outline some of the decision making on the design and potential data sources for this study while also factoring in how the study approach made sure to collect qualitative data and triangulate it as well.

Teacher impressions. Of Smith's (1993) five data sources, none was more important for this study than teacher impressions. Smith (1993) claimed that this data source "may be the single best measure" of adequacy (p. 157) and that it offered the "most powerful method for assessing adequacy" (p. 197) because such perception was sensitive enough to distinguish even those students who fall between courses. Smith (1993) made it clear that such perception was so sensitive in part because it was based on more than one piece of writing and it contained "knowledge of the student as a student and as a person" (p. 157). Since it was postulated that Cardinal College operates on Smith's expert-reader model—which hinges on the teaching perception of placement readers to make placement decisions—it was only natural that, to check the adequacy of that placement decision in this case, such teacher impression data would have to be collected from the teachers of the courses in which students were placed.

Additionally, teacher impression data was a logical check on the interpretive argument we have outlined at Cardinal College. The goal of validation is to check the interpretive argument, and our interpretive argument claims that expert teacher placement readers are making course-related decisions. Therefore, we have a responsibility to check that assumption, and we could use the impressions of course teachers and a qualitative description of the reading and writing criteria they discussed in students' essays as they

checked course placement and worked with students' papers in the class to test our assumption.

While Smith's (1993) acknowledgement of the value of teacher perception as a data source was positive, he also admitted potential problems. In particular, he raised issues with timing, or when to collect this data. He discussed how teacher perception "changes considerably across the term" as teachers have "too little evidence" early on in the first weeks, but past midterm, their perceptions are correlated with the final course goals and grade more than with the determination of adequate placement (Smith, 1993, pp. 157-158). Therefore, Smith (1993) asserted that "teacher perception data must be collected somewhere near the end of the first third of the course, between week 3 and week 5" of a 15-week term (p. 158).

The plan for getting teachers' impressions. As I planned the study to use teacher perception data, I chose the timing to fall within or as close as possible to Smith's suggested range. To be more precise on the design, this study was planned to collect teachers' perceptions of adequacy in a number of ways, both directly and indirectly. First, I would directly collect teachers' perceptions about the appropriateness of placement as I watched teachers read and think-aloud about course diagnostic exams (administered, in part, to check placement during the first week of class). This meeting would take place during week two or three of the semester (based on teacher availability). Given the potential flaws with making a placement decision on only one piece of writing, teachers at Cardinal College are required to administer a diagnostic to collect an additional piece of writing to check placement. Therefore, one of the key reading goals

teachers have during the reading of the diagnostic is to check the adequacy of placement. I would glimpse that directly as they naturally made such determinations.

A second, less direct way the study was planned to collect teachers' perceptions on the adequacy of placement was watching teachers read and think-aloud about one course paper collected as close as possible to week four. The meeting with teachers was planned to take place between weeks four and six of the semester (again based on teacher availability and the due date of their first major course paper). Here, the teachers would not focus directly on determining the adequacy of placement, but several aspects of what they discussed would be important: Had they refined or changed their assessment of specific students or the overall adequacy of placement from when they first discussed their diagnostic exams? For instance, did they feel a particular student was stronger now than they did with just the diagnostic exam? Such information would complement teachers' earlier assessment of adequacy, as it was based on more evidence about each student, as well as a piece of writing that was not written on-demand.

Additionally, by looking across the data—what all Course 10 and then all Course 11 teachers shared in terms of course criteria, where they agreed about the relative importance of certain course goals, etc.—this study would begin to compile a sense of the course criteria and goals for both Course 10 and Course 11. The power of these criteria and goals would be that they emerged directly from the natural work of teachers of the courses. And, such an understanding of course criteria could be an indirect way of getting teachers' perception of the adequacy of placement when such criteria were compared to the criteria under discussion during placement reading. Therefore, I would also watch placement readers read and think aloud about placement essays in order to compile a

similar understanding of the criteria under discussion during placement into Course 10 and placement into Course 11.

Using think-aloud protocols. As I have already discussed, think aloud protocols would be one way that teachers' perceptions of adequacy would be gathered for this study, and they would be collected as teachers read and thought-aloud about students' diagnostic essays and, additionally, at least one piece of students' course work. Therefore, some additional theory on think-aloud protocols sheds light on the study design.

Think-aloud protocols are one kind of the larger data type that Ericsson and Simon (1980) have termed “verbal reports” (p. 215). In Ericsson and Simon's (1980) article, as well as their more comprehensive work (1984), Ericsson and Simon distinguished between verbal data collected during the completion of a process or task—*concurrent*—and that collected afterwards—*retrospective* (p. 16). Think-aloud protocols fall under concurrent verbal data. While a research subject completes a task, problem, or process, they are asked to think-aloud, essentially to speak their thoughts out loud.

As Ericsson and Simon (1980, 1984) discussed verbal reports, they outlined a number of issues that have been raised historically on the validity and reliability of such data. They detailed a strong negative view of verbal reports in the literature that claimed verbal reports were merely introspection—that is, they may not truly reflect actual mental processes or thoughts and therefore would need to be verified by other data (Nisbett & Wilson, 1977). Through a comprehensive review and analysis of research studies, Ericsson and Simon (1980) concluded that many of such concerns about verbal reports could be put to rest once it was recognized that many problematic studies did not follow

systematic collection guidelines consistent with what was known about how short-term memory and long-term memory work within a processing model.

To give some examples of their systematic collection advice, I offer two guidelines. First, Ericsson and Simon (1980) argued that more general directions for concurrent verbal report participants—where they are instructed only to think-aloud during a task—would be better because such instructions would interfere less with the natural task and the process participants take. Second, they argued that ensuring the information of interest in the study is actually the information that is attended to by the participants directly in the task will reduce interference in concurrent participants' reports, as well as ensure that retrospective study participants are pulling from memory of how they actually completed the task instead of from some theory or general understanding. While Ericsson and Simon (1980, 1984) concluded that verbal reports could be trustworthy data if systematically collected and controlled, they also discussed how such reports did increase the time it took to complete the task, and participants could produce very little verbalization if the task was so familiar as to become automated.

Ericsson and Simon's (1980, 1984) work had a great effect on the acceptability of verbal reports as psychological data, and additionally, verbal reports became of interest to those doing composition research who were interested in the mental processes involved in writing. The use of protocol analysis in composition research has informed such studies as those looking at the composing process of writers (Flower & Hayes, 1981; Hayes & Flower, 1983; Swarts, Flower, & Hayes, 1984; Witte, 1987) and even how student writers respond to the comments of their teachers (Hayes & Daiker, 1984). One area in the literature of interest to this study is the use of verbal reports—think-aloud

protocols specifically—in a way quite different from these previously mentioned studies looking at the composing processes of students.

Smith's (1992/2009) influential discovery on the value of placement readers most-recent course teaching expertise to the adequacy of their placement decisions could not have come about without the use of think-aloud protocols that revealed that placement raters who had recent experience teaching the course into which they believed a student should be placed would often discuss more than writing criteria; they would discuss "the writer as a student," noting how the student would likely behave in class or where they would struggle or excel (p. 183). And, Smith's larger expert-reader model, which would not have been possible without the understanding that the think-aloud protocols revealed, was in part influenced by the growing interest in such data and Huot's (1988) use of think-aloud protocols in his dissertation on which Smith served as a non-official reader (Huot, 2007, pp. 18-19). Huot (1988) examined the work of rating student essays using think-aloud protocols (as did his 1993 piece), and think-aloud protocols helped him uncover and examine what raters discussed and the differences between novice and expert raters.

Given the usefulness of think-aloud protocols to Smith's (1992/2009, 1993) work, the current study would seek to use the same data to investigate an assumption of Cardinal College's placement program that had been built on Smith's work—namely that placement readers are using course expertise to make adequate placement decisions. The use of think-aloud protocols for assessing the reading of student work seemed a worthwhile endeavor particularly given Huot's (1993) push for using such verbal report

data to investigate “how teachers and raters make evaluative decisions about student writing” (p. 228). He wrote,

The most important information to come from the present study as it relates to future research is the use of verbal data collection for more insight into the rating process....There is little doubt that the successful use of protocol analysis in the present study should provide an opportunity for future research studies into the process of rating student writing and the factors which most affect this process.
(Huot, 1993, p. 228)

Huot (1993) went on to suggest that the effect of “purpose and context” on teachers’ assessment criteria will be particularly interesting (p. 228). Building on these arguments, this study was planned to use protocol analysis to investigate the reading of student work in a very specific purpose and context—that is, placement reading. And given that placement readers were theorized to use course criteria to determine placement, protocol analysis was used to investigate teachers’ process of determining placement adequacy and reading student work during the early part of the term.

My rationale for the use of think-aloud protocols also had a basic logical element as well. Hayes and Flower (1983) discussed think-aloud protocols focused more on investigations into the composing process, yet their explanation of the value of “process-tracing methods” of data collection (of which think-aloud protocols are one kind) was useful the focus of this study—the reading and evaluation of student work (p. 211). Hayes and Flower emphasized that there is often much that is going on during a task that is not viewable in the finished product. This is certainly true of reading student essays. Looking only at what teachers choose to comment on in written comments or what

decisions teachers make after reading (as in placement or exit exam readings) can only tell an incomplete picture. The use of think-aloud protocols could reveal some picture of that behind the scenes process. While Ericsson and Simon (1980, 1984) cautioned that what is revealed will never be a complete record of the thoughts, think-aloud protocols could provide some data that would help Cardinal College study its placement process.

Since training and the directions specifically given to subjects as they provide think-aloud data are essential pieces of Ericsson and Simon's (1980, 1984) rebuttal of the critique of verbal reports as data, I want to make it clear that this study was planned to provide general directions to participants—both placement readers and course teachers—so as to avoid disrupting their natural process of reading as much as was possible. I cannot claim that it is possible to totally limit disruption since just the act of studying can cause participants, who are aware of being studied, to act differently than they might otherwise—potentially looking for ways to be more agreeable and cooperative or even causing them to be more guarded (Maxwell, 1996). However, the general directions would ensure that participants were not led to particular observations. Ericsson and Simon (1984) also suggested that think-aloud sessions be conducted with the researcher out of the view or behind the participant so that subjects do not address comments to the researcher but complete the task as they naturally would (pp. 375-377). Given space constraints at Cardinal College, it would not be possible to arrange such an environment. However, I planned to be careful to monitor such social comments, keeping track of participants who had problems with such comments when reviewing transcripts so I could discuss this issue in relation to the data collected.

Triangulation through teacher interviews and document analysis. Despite the value and strength I have established for think-aloud protocols, one tenet of qualitative research—that establishes its credibility—is triangulation (Lincoln & Guba, 1985). By triangulating data—collecting multiple kinds of data or multiple instances within one kind of data that point to the same information—researchers can use one set of data to corroborate another. Additionally, and perhaps more importantly, multiple sets of data can add nuance and depth to our investigation. Therefore, in addition to teachers’ think-aloud protocols, this study was planned to also interview placement readers and interview teachers through a semi-structured approach (Denzin & Lincoln, 2008). In the interviews, I planned to ask readers and teachers to explain how they understood the courses and program, how they approached teaching, how they felt about the placement prompt, and what they felt were the writing strengths and issues characteristic of the students in their course of expertise. While I designed questions, I would also allow the participants’ responses to spark new questions or avenues of information. In addition to the semi-structured interviews, the study was planned to also allow for additional interviews, if necessary. As earlier data was analyzed and coded, if questions or concerns arose for placement readers or course teachers, the study would add follow-up interviews to subsequent interactions or at the completion of the study.

As a further data source for triangulation, teachers’ course materials—syllabi, assignments, rubrics, etc.—would also be collected. These documents of the teachers’ courses would be analyzed and connected with what they said in interviews and discussed in think-aloud protocols as the important reading and writing criteria of the course. These additional data sets—teacher interviews and course documents—would

provide another set of data on the criteria teachers deemed important for their particular course in the writing program (and a beginning sense of the effectiveness of the placement prompt) which could be triangulated with think-aloud protocols.

Number of students moved to a different course at the beginning of the semester. While it would seem a direct measure of adequacy, there are some factors to consider with using the number of students that teachers elect to move. Smith (1993) explained that the number of students moved during the first weeks because they are misplaced is a potentially problematic data source on the adequacy of placement since course teachers may be reluctant to move students for a number of reasons (p. 156). First, moving students would open spaces in the roster for other students to add the course late, and teachers may avoid the burden of catching up new students by not moving students out of the class. Second, teachers of lower courses may want to keep students who are stronger writers in order to have those students as potential leaders in their class. And third, when teachers determine that some students are weaker writers in the course, they may keep these students instead of pushing them down because they believe in their own ability to catch the students up. (Here Smith noted that, in his experience, his teachers were not wrong about such claims to their ability.)

Yet, Smith (1993) suggested that, even with the potential flaws, this quantitative data should be tabulated each term. In his mind, having such a record of numbers allows a specific term to be compared to an established “norm” and gives one set of data that can be triangulated to other sets of data on the “adequacy” of any particular placement term (Smith, 1993, p. 158).

Determining the number moved (and adding to teachers' impressions) through an adequacy survey. Given these concerns, the study planned to collect this number, but to do so in a way that would also tap teachers' impressions. As teachers read and thought-aloud about the diagnostic essays and course papers, they would be asked to complete a survey tallying the number of students who were *prototypic of the course*, *slightly below*, or *slightly above*, as well as those students who teachers felt *should have been placed in higher course* or *should have been placed in lower course*. In this way, the study would have an official number (the survey at the diagnostic stage would offer the official number of students that teachers were recommending to be moved) and an interesting facet would be added to the teacher perception data at the same time, allowing for comparisons across the participants and data sets.

Final course grades. Smith (1993) offered a word of caution about the usefulness of final course grades as adequacy data. Since not all teachers have the same criteria for determining specific letter grades, and since "studenting" issues can affect final grades as well as writing issues, this data source could be influenced by these other issues, potentially making what is learned from final grades unclear (Smith, 1993, p. 157). Smith (1993) suggested determining median grades and sorting students "above the median, on the median, and below the median" to correct for some of these issues (p. 157).

Collecting final course grades (and adding to teachers' impressions) through an end of term review sheet. At the end of the semester, final course grades would be collected, and the study would sort the students around median grades to limit differences across teachers. Yet additionally, the study planned to collect the final course grades in a way that offered additional information on teachers' impressions and criteria. Teachers

would be asked to complete a sheet that listed students' final grades, but that also had space for a brief description of the teachers' final assessments on students' portfolios. Such information would allow the study to connect earlier teacher impressions with those at the end of the term. For example, as teachers discussed student work and completed the survey for their course, if they indicated a number of students were slightly below the course, collecting the final course grades in this way would allow another check on the teachers' impressions of adequacy. Were the final grades lower overall for these less than prototypic students? Or, did the final grades suggest that some were able to handle the course?

Exit exams. Having exit exam data is potentially interesting since Smith (1993) explained that exit exams can often be more useful than final course grades since just the writing is considered and not studenting issues. Further, he explained that exit exams can provide writing programs a way to compare lower course students exit essays with the placement essays of those directly placed into the higher course. Such information can help a program see that students placed in a lower course are being well prepared for a higher course.

No exit exam at Cardinal College. Since there is no established exit exam at Cardinal College (although some teachers administer a final exam of their own choice), this study did not collect exit exam data. However, it is possible that future studies could consider exit exams should they be added to Course 10. To have some sense of the exit abilities of students though, copies were made of students' final course portfolios. It was thought that such information may prove useful to compare with students' earlier work.

Students' perceptions. The final piece of data that Smith (1993) discussed on checking the adequacy of placement was students' impressions. For Smith, students' impressions were potentially problematic because, if collected during the course, they may be influenced by the common belief of lower-course students that they do not belong in the lower course. However, as Smith discussed conversations he had with lower-course students semesters later, he revealed that many come to see the value of the lower course.

Students' perceptions outside the scope of this study. It was not in the scope and timeline of this study to collect students' impressions as the focus here was on checking one key assumptions of our interpretive argument that related to teacher expertise; however, later studies of Cardinal College's placement process should be planned to collect students' impressions. Specifically, it would be interesting to follow up with the students that participated in this study after they have completed the writing program to discuss their impressions.

Relevant Literature Influencing Decisions on Coding

It is important to stress that as I coded and analyzed data, I planned to follow the suggestion of Creswell (2007) to let the code emerge from the data as opposed to using a preexisting code: When the code emerges from the data, it can reflect the views of participants inside the research more fully since it is not limited to previously determined units (pp. 152-153). Therefore, I would code each think-aloud session and interview for themes and "significant statements," and group them "into meaningful units" (Creswell, 2007, p. 156). Given the research question of this study to investigate the assumption that we place using course criteria, I would code to determine specific criteria of reading, writing, and thinking under discussion for placement in Course 10 and Course 11.

Additionally though, since Smith (1992/2009) specifically discussed that expert placement readers may also note behaviors of “the writer as a student” as opposed to just the characteristics of the essay—which Smith claims is *privileged course knowledge* that helps expert readers make their placement decisions—I would also develop codes for these statements (p. 183).

Even though I would allow the code to emerge from the data, I was already aware of other codes which are important to mention as they would likely influence my thinking during the data analysis. Huot (1988) specifically developed a code that relates to essay raters which noted such things as positive, negative, and neutral comments, as well as comments directed toward specific writing attributes noted by raters such as content, organization, tone, style, etc. (p. 113). On this second part, Huot based his code on a modification of the now famous Diederich, French, and Carlton (1961) code which sought to uncover specific factors on which raters could agree that influenced ratings. In fact, Huot (1990b) specifically discussed how Diederich, French, and Carlton’s work is a key reference point for how we read essays (p. 251). However, Huot (1988) found that their code did not work fully for his data, so he modified it to fit his context, purpose, and data—making sure to develop a system that allowed him to fully code his transcripts.

Additionally, given the particular context and purpose of Cardinal College’s courses, concerns of interest at Cardinal College would likely emerge in the data that would not be represented by existing codes. For instance, the emphasis the College places on academic reading is not something covered in either Huot or Diederich, French, and Carlton’s codes. Therefore, such data would dictate that I develop relevant code. On a final note, I want to stress that, common to qualitative research, the plan for analysis and

coding of this study included a check on the clarity of the code through the sharing of samples of coded transcripts with at least two colleagues who have a background in writing but were not associated with the study. By asking them to discuss their observations on the logic of the coding system with me, I could become aware of my own assumptions and more fairly understand the vision and viewpoint of participants.

Lining Up Decisions with Common Evidence of Validation

Both the AERA, APA, and NCME (1999) standards and Kane (2006) discussed how empirical studies designed to evaluate the most significant or fundamental interpretations involved in a test's use can allow such interpretations to be tested and refined. The requirements for gathering evidence in support of the interpretive argument can quickly become overwhelming without some way to prioritize. As I discussed earlier, the interpretive argument itself provides the most significant way to determine where to put resources and energy in validation studies. If pieces of our interpretive argument are in question or uncertain, we must choose to put out time, money, and energy to gathering the kind of evidence that will help us evaluate those key assumptions. In fact, Cronbach (1989) argued that we can decide what to study in validation by considering "prior uncertainty," "information yield," "cost" and "leverage" (p. 165). In short, Cronbach claimed that we should study those aspects that are in question, those that we can get good data on, those that are relatively affordable, and those that can be useful to presenting our case to others with differing viewpoints.

Given that we must put our time, energy, and money into studies that will help us evaluate our particular interpretive argument, and that we must weigh which studies will provide us with the best balance of usefulness, information, and cost, the planning for this

study has emphasized the specifics of Cardinal College's case and context. However, the AERA, APA, and NCME (1999) standards outline five kinds of evidence for validation, so I want to compare the plan for this study against the suggested evidence to show that while this plan emphasizes qualitative data and teachers' conceptual frameworks, it has still planned data sources that will uncover two of the kinds of evidence discussed by the standards directly and that will provide potential information on two of the other kinds of evidence as well. The final type of evidence recommended by the standards is immaterial to Cardinal College's context.

Before I proceed with the five kinds of evidence outlined by the standards, it is interesting to note that the standards began their discussion of these kinds of evidence by clarifying first that "validity is a unitary concept" (AERA, APA, & NCME, 1999, p. 11). The AERA, APA, and NCME felt compelled to make this clarification precisely because the kinds of evidence they outlined echo concepts once thought of as distinct and separate kinds of validity, as I described in Chapter Two. Thus, they were careful to make sure readers could see that while particular kinds of evidence can involve the content, construct, or criteria of a measure, such evidence alone was not the whole of validity for, as I've already discussed, validity is an ongoing and multi-lined inquiry into the adequacy of a test's use.

The first kind of evidence that the AERA, APA, and NCME (1999) standards discussed was "evidence based on test content" which is "an analysis of the relationship between a test's content and the construct it is intended to measure" (p. 11). Studies gathering such evidence look at the "themes wording and format of the items, tasks, or questions on a test, as well as the guidelines for procedures regarding administration and

scoring” (AERA, APA, & NCME, 1999, p. 11). The key issue is how well the particulars of the test represent “the content domain,” as well as how that domain relates to the construct specified by the interpretive argument, and such determinations are often based on the judgment of experts (AERA, APA & NCME, 1999, p. 11). This kind of evidence, while important, was only partially planned into the study. Those portions of test content related to scoring were definitely a part of the data planned for the study, as placement readers were studied. But, those portions of the description of test content which relate more to looking at the wording of the test and guidelines (essentially, analyzing the placement prompt), were not directly looked at in this study. Future studies would have to look more closely at this kind of evidence. Or, since the data collected for this study included the placement prompt and copies of students’ responses, as well as teachers’ understandings and impressions of the test, the data collected here could be reanalyzed to offer a beginning look at test content in the future.

The second kind of evidence was “evidence based on response processes” (AERA, APA, & NCME, 1999, p. 12). These processes can look at how examinees respond or how judges or scorers respond, and the goal is to check if such responses are consistent with the assumptions of the interpretive argument (p. 12). On the examinee end, such evidence is often gathered by asking test takers about “their performance strategies or responses to particular items” or gathering other kinds of evidence on such processes such as drafts or “electronically monitored” information (AERA, APA, & NCME, 1999, p. 12). And the evidence is similar for judges/scorers who “evaluate examinees’ performances or products,” and such judges can be asked about their processes or studied in some other way (AERA, APA, & NCME, p. 12). The purpose

with judges though is to check that “if [they] are expected to apply particular criteria in scoring examinees’ performances” that “they are, in fact, applying the appropriate criteria and not being influenced by factors that are irrelevant to the intended interpretation” (AERA, APA, & NCME, 1999, p. 13). This type of evidence spoke directly to the focus of this study, which posed a question about placement readers’ criteria. Thus, types of evidence were chosen that would get at the process of placement readers and their particular criteria. In short, the plan for this study directly collects data on the response processes of placement readers.

The third kind of evidence was “evidence based on internal structure” which relates to multi-item tests in particular and if the items on the test relate in ways that the interpretive argument and particular uses assume (AERA, APA, & NCME, 1999, p. 13). The standards provided a few examples which were helpful for unpacking a type of evidence that can vary greatly for across different uses of a test. One example they discussed has to do with difficulty of test items. If, in a particular use, certain items on a test are assumed to be more difficult than others, relevant evidence in this category would involve if test performances seem to bear out that assumption (p. 13). This evidence had little bearing on the context of Cardinal College, as Cardinal College’s placement test is not a multi-item test.

The fourth kind of evidence was “evidence based on relations to other variables” (AERA, APA, & NCME, 1999, p. 13). This kind of evidence looks at the relationship of the test score to other measures, performances, or situations in order to evaluate if it conforms to the assumptions of the interpretive argument. As with all the kinds of evidence described here, the particulars of the argument will affect what must be

examined. So if the interpretive argument suggests that the test under study will produce scores that “diverge” from those on other existing tests, then evidence should align with this interpretation (AERA, APA, & NCME, 1999, p. 13). Based on particular interpretive arguments, how well a test predicts a future performance may be important. Therefore, evidence of this kind could show a relationship between the test score and other current or future measures or performances. In the case of this study, data on course performance, such as course grades and final course assessments, were directly collected. Therefore, the relationship of the placement decision to other measures was directly considered.

The fifth kind of evidence was “evidence based on the consequences of testing” which looks at “the intended and unintended consequences of test use” as part of determining validity (AERA, APA, & NCME, 1999, p. 16). In the case of Cardinal College’s interpretive argument, final grades could be one form of evidence that might reveal the consequences of placement decisions. Yet the issue is much larger than that. Some potentially negative consequences cannot be seen through grades alone. It is possible for a student to pass but to have had the course experience demotivate them or turn the off from future learning on the subject. Further, the discussion of consequences by the standards was not only for negative or unintended outcomes. The standards also made clear that even intended positive outcomes would need to be documented. So, again, while the grade might be passing, if, for instance, a test’s use was claimed to have positive effects on teaching and learning, evidence would need to support such an assertion. On this final kind of evidence, Cardinal College’s planned data might offer

potential implications on the consequence of testing, but future studies would have to look at this more directly.

Summary of Study Approach, Data, and Analysis Plan

The study was planned to be primarily qualitative in design (interviews, think-aloud protocols, surveys, and the collection of documents). The goal was to closely investigate the particular context of Cardinal College's placement process, so the particulars of this writing program were detailed and, through in-depth qualitative data collection, the research question was answered. However, some quantitative data was also planned—such as the number of students placed into various courses, the number receiving particular final grades, the number felt to be inadequately placed, and frequency counts of particular codes or themes that emerged during analysis. Such quantitative data would be used to support the understanding and interpretation of the qualitative data.

In holding to a qualitative research approach, this study conceived of issues of the validity, credibility, reliability, and dependability of the study in more qualitative terms. Thus, the study was not planned thinking of these concepts in the classic positivistic stance (where objective truth is the goal), but, as Creswell (2007) described, the study was designed to make the research valid, reliable, credible, and transferable in terms of what it means to do good qualitative research (where the goal is understanding based in a constructivist sense). Therefore, as Creswell specified, I have provided thick descriptions of the context, situation, site, and participants of this study to aid those with similar situations in transferring understandings; made clear the reasons/argument for procedures; planned for prolonged engagement with the participants; factored in the collection of multiple sources of data; outlined my own experiences and positions;

planned procedures so as not to lead or directly influence participants responses (in interviews or think-alouds) and to ensure the accuracy of transcripts and codes.

In this study, the assumption to be investigated asked if expert placement readers at Cardinal College were using course teaching and learning expertise to make placement decisions. To get at this information, the study was planned so that data was collected both from placement readers and course teachers. Specifically, the plan included interviews and the collection of think-aloud protocols (Ericsson & Simon, 1984), as well as placement and course documents, to investigate the reading, writing, and thinking criteria under discussion by expert teacher-readers at placement and the criteria under discussion by teachers in first-year courses. In addition, teachers would be asked to assess the adequacy of placement through a survey, and students' final grades would be collected.

By comparing the criteria uncovered from these various pieces of data, the study was planned to uncover the relationship between the criteria of placement and the criteria of the course. Is teachers' course expertise indeed being utilized during placement, as our program postulates? Exploring these criteria would reveal if our way of placement offers a pedagogically useful and purposeful view of student's placement essays. In short, uncovering these criteria and examining teacher expertise would aid Cardinal College in the ongoing study of its program and in crafting a validity argument for its placement process and procedures by helping us check this one assumption on which our practice rests.

Steps of Data Collection

Now that I have provided a rationale for the data collection and analysis plan, I want to move to describe the actual procedures of the study, as they happened. In what follows, I outline the data collection procedure for this study chronologically in step-by-step detail through three phases of data collection: phase one at placement, phase two during students' first-semester course, and phase three at the end of the semester when placement was reassessed and the final course results and grades were collected. I also offer a final note on confidentiality overall and the process of debriefing.

As I describe the data collection steps, I also indicate where the data collection procedures may have changed from initial plans for the study. Some changes were made to the initial plan because as I collected data, I allowed interactions with participants, participants' needs, and the analysis of earlier data collected to alter the design—an idea that coincides with the idea of “emergent design” (Creswell, 2007, p. 38). In short, later research tasks and plans benefitted from and reflected experience and knowledge gained in earlier phases. Since no significant changes were made to the research plan, the study did not have to go through a second review by the Institutional Review Board (IRB) at Indiana University of Pennsylvania (IUP) or at Cardinal College.

Phase One: Placement

The first phase of data collection began during placement testing the summer prior to the start of the Fall 2013 semester. This phase of data collection focused on securing the participation of two placement teacher-readers—one responsible for reading essays as an expert for each of the two courses—and on securing the participation of at least 50 students who would agree to allow their placement essays to be studied. The goal

in this phase of study was to seek placement readers and students' placement essays to determine the criteria under discussion during placement for each course so I could determine how placement criteria align with course criteria.

As outlined above, the data collected during this phase of the study included placement readers think aloud protocols, copies of the placement prompt and readings, copies of student essays, and reports on final counts of numbers placed overall and in each course. In what follows, I offer background on how the placement sessions to study were selected, and I describe a backup plan for the study if less than 50 students consented during placement. Next, I describe the procedures for obtaining readers' and students' informed consent, and I indicate the number of participants secured. Finally, I detail the steps of data collection, including what steps I took prior to the arrival of readers, during the sessions, and following the sessions.

Selecting placement dates. Some background on the placement testing dates and the selection of which dates to study is necessary. There were six scheduled placement dates, published in January 2013: April 13, May 31, June 22, July 27, August 9, and August 15. Additionally, there was an additional late-admit student session added for August 22, bringing the total number of sessions to seven. The number of sessions was a reduction from previous years where there were ten prescheduled placement sessions and additional sessions added as needed for late-admit students.

The office responsible for testing at the College reduced the number of prescheduled sessions because it found that many sessions were poorly attended. Instead of having two sessions per month from April to August that were lightly attended, the office preferred to fill more seats for one larger session per month. An exception was

made for August, and two sessions were still maintained for that month, as well as an additional session added. The office made this decision because in previous years, despite the number of sessions offered, there has been a fairly consistent population of students who make their decision to attend late in the summer or who arrive late (on international visas). Many of these students never attend a summer placement session, and instead they require placement testing the week prior to the start of classes, or occasionally, they show up during the first week of classes and must take placement tests before enrolling. The office of testing felt that if they did away with the dual August sessions, the number of students who remained unregistered at the start of the semester and still required placement testing would increase, subsequently disrupting classes during the first and second weeks.

While there were seven placement sessions total, it was decided that this study would begin collection during the final three placement sessions. This decision was made for several reasons. First, previous number breakdowns along with the testing office's reduction of sessions supported the idea that a significant number of student placements could be read, even focusing on later sessions. During placement testing for 2011, the majority of students took their placement in June or after (64%), and a fair number of students took their placement during the August placement sessions (15% in 2011). Since the placement office reduced the total number of placement sessions, and since seats were limited to approximately 50 testers at each session, attendance at the final sessions was expected to increase, allowing for a good number of student placements to be studied.

Second, during placement testing for several of the years prior to Fall 2013, earlier sessions had a higher number of students placed in Course 11 with fewer students

placing into Course 10 than in later sessions. It seems that in many of the previous years, students who were well prepared for college level work also preferred to make their attendance decisions earlier. Thus, collecting data on the criteria discussed for each course was expected to be more balanced in later sessions when there were likely to be more potential Course 10 students.

Third, and perhaps the most important reason why data was not collected at all sessions, Cardinal College's writing program recently changed their placement procedure to include two expert readers for each session (one for Course 10 and one for Course 11) as opposed to the earlier practice of just a Course 10 reader (a change initiated to check for tweeners). Therefore, it was important to the program for the first sessions to run unstudied as readers acclimated to the new process. Instead, beginning with the August 9th session and continuing through the August 15th and August 22nd sessions, student consent would be sought from students taking the placement test. However, even though direct student participants would only be sought in the final three sessions, final counts on the placed students would be collected from the writing program director for each of the earlier sessions and overall counts for all of summer placement once all placement testing was complete. These counts would include the total number of students placed into Course 10, the total number placed in Course 11, and the total number of tweeners.

Backup plan. It was not certain, given these changes and restrictions, that I would secure a sufficient number of students' consent at placement. The goal was to get at least 50 placement students (sufficiently balanced in each course) to participate. If less than 50 students allowed their placement essays to be studied, another option was planned to secure enough placement essays to determine the criteria of placement. As

classes started the week following the final placement sessions, it was possible to ask students who consented to participate in the second phase of study (during the first-semester course) to also grant me permission to pull their placement essays and records (including any notes indicated by placement readers on their decisions and audio recordings of readers' thinking and discussion during the last three placement sessions). If there were not enough students from the last three sessions now consenting or the placement notes on earlier sessions were not sufficient enough to provide specific criteria on the required number of students, then I planned to meet with placement readers during phase three of the study to have them reassess the essays of all students who consented during the course phase. The re-reading of essays would, of course, be blind; placement readers would not know their initial decision. This would allow me to gather specific criteria relevant to a placement decision while I also checked if the readers would make the same decision on the same essay. ²

Informed consent and number of participants. Before the first session, I initiated contact with the two teacher-readers who were participating in placement this summer (one from Course 10 and one from Course 11) by emailing them about the study (see Appendix B), and I attached an informed consent form to the email (see Appendix C) asking to meet with them to discuss the study and the form. The form explained the study's purpose and value—to assist Cardinal College by collecting validation evidence for the placement process at the College—as well as its value in helping the field develop and understand local assessments. The consent form also explained that readers were consenting to read placement essays during the session by voicing out loud their thinking

² More is said about the backup plan for securing enough placement criteria data and its procedures later in this description of the phases of study as relevant steps of the backup plan became necessary.

process as they read each essay, and that they had the right to consent or not, as well as the right to withdraw at any time. It was explained that outside of an increase in reading time, since thinking will be voiced and recorded instead of noted silently, there were no known risks for teacher-readers. There were potential benefits though since study of the criteria under discussion at placement could help teachers make their reading of student work more conscious and explicit. Two placement readers—Adele who read for Course 10 and Theresa who read for Course 11—participated in think-aloud protocols as they read and placed student essays in these two courses.

Additionally, as I attended the three final placement-testing sessions at the site, I met with students at the beginning of the testing session directly. I distributed two informed consent forms to each student at the start of placement-testing sessions, so they could keep one for their records if they chose to participate (see Appendix D). I discussed the study and the form with the students, answering any questions as necessary. Students were instructed to place their completed or blank form (depending on their choice to consent or not) in the same box upon exit from the class so that no one in the session would know if students chose to participate or not. Since this study examined a normal testing process already in practice at the institution, there were no risks or discomforts for students during this phase of the research that were not normally associated with existing practices, and their consent form explained this as well.

Despite the lack of risks, zero students consented from the August 9th session out of 17 students, one student consented from the August 15th session out of 16 students, and zero students consented from the August 22nd session out of 12 students. In retrospect, I attribute this low number of initially consenting participants to the awkward and

apprehensive state students find themselves in during Cardinal College's placement testing. When students at any college or university come to a placement session, they are likely apprehensive and unsure. Additionally though, when students arrive at Cardinal College's placement sessions, many of them are joined by their parents who can attend financial aid and parent orientation presentations while students take their tests. These students are used to seeing their parents as the authority, so much so that despite not being in the immediate vicinity of their parents while taking their tests, they may have been reluctant to agree to anything without consulting their parents.

In addition, since the office of the college registrar seems to have, just before the testing session, talked about FERPA and students' right to grant their parents permission to access their educational records or not, it may be that I asked students to consent to participate at just the wrong time in terms of how they saw their own capacity for autonomous decision making. When parents realized that their students had to grant them access to records even though they were paying the bill, many parents turned to their students to assert their authority and insist the student take whatever steps to grant them access (a fact that has since caused the office of the registrar to take steps to rethink when and how they discuss FERPA rights with students). I believe my informed consent session, which began once students were relocated to a testing room just after the FERPA session, and my explanation of the study and its lack of risks could not be heard over students' worry about losing their parents' approval and support. I believe so few students consented because they were more worried about their parents' feelings than they were about the actual study and what it entailed, and they chose to take the arguably

safer route, in terms of angering their parents, by not signing anything their parents had not given them permission to sign.

Having only one student agree to consent prior to the reading sessions meant that the backup plan was enacted during phases two and three. Despite being collected later, the analysis of the data collected as part of the backup plan would impact the understanding of the placement criteria. Therefore, it is important to note here that early in phase two, 118 students in total agreed to allow me to pull their original placement notes, recordings, and placement essays. Accessing students' initial recordings, as they were still being transcribed and were not yet destroyed, changed the total number of students granting permission for use of their original placement records. Nine students from the August 9th session (out of 17 students), 11 students, including the initially-consented student from the August 15th session (out of 16 students), and six students from the August 22nd session (out of 12 students) were now consenting to have their original placement decision and essay studied.

The inclusion of the backup plan students brought the total number of students consenting with original placement sessions records to 26—12 placed in Course 10 and 14 placed in Course 11—with one of these students being designated a tweener, and eight students having notes next to their names signaling they would have been placed in the old Course 11A class. Additionally, the group was made up of 53.85% female students and 46.15% male students with 11.54% of the total being international students. The make-up of the group was not that dissimilar from the population of first-time, full time freshmen overall, having about 3% more male students and consequently 3% fewer

female students than the population of first-time, full time students and almost 5.5% fewer ESL students.

The number now consenting also confirms my sense that something other than the objection to the study was part of what impacted students' willingness to consent earlier since 56.82% (25/44) of students who had not consented when first asked now did consent. Further, these same students, as well as others representing 48.76% (118/242) of the overall total number of students enrolled in first-semester courses, granted consent for phase two when that phase by comparison would likely seem more invasive as it involved far more pieces of the students' writing and even ongoing interaction with their teacher.

Reading session procedures prior to reader arrival. After the student essays were written, the readers typically read and placed students immediately, meeting directly after on the same day as the placement testing session. The only exception to this was the August 15th session, where readers met three days later to read. I followed the same procedure each time. As I had to seek consent from students prior, I was already on campus. Thus, before the arrival of readers, I collected the placement essays from the office for testing. Not wanting the readers to know which essays were participating and which were not, I copied and created an identification number for every essay, recording the number on the top of the essay as well as in a log of notes for the day. Also prior to the first session, I obtained a copy of the exact writing placement prompt as it was handed out this year. The prompt was analyzed later to understand the task demands of the prompt (along with course materials that were collected during phase two), but at this stage of research, I wanted to have the prompt for use during sessions.

Procedure during reading sessions. Initially, the plan was for the first placement session to begin with an interview of each teacher-reader on her most recent course teaching experience, her background, her experience, and her process of placement reading (see attached placement interview questions in Appendix E for planned questions). However, since both readers had to finish reading as soon as possible that day, I decided to postpone the interview until later. We ended up completing the interview during phase two since (a) both readers had already agreed to participate during phase two as teachers, (b) most of the interview questions already planned for course teachers during that phase were identical to the planned placement interview questions, and (c) the only question that was different—asking about their placement reading process—would already be uncovered through analysis of what they said as they read the essays and would end up being answered directly in comments readers naturally made about their reading process as they thought-aloud about student essays.

I followed the same basic steps for the three original placement reading sessions: (a) I conducted and audio-recorded think-aloud protocol training with the Course 11 reader (see Appendix F) at the start of the first session, but for subsequent sessions, I extended the option to the reader to decide if they wanted to revisit training. (b) I collected and audio-recorded open-ended think-aloud sessions as the Course 11 reader read all student essays and made decisions about acceptance or rejection from Course 11. I allowed the reader to notice what she would and describe what she saw as she normally would. The only prompting was in response to sustained silences, reminding the reader to think-aloud. (c) During the think-aloud sessions, I logged each essay and placement decision so it was clear what order essays were discussed in on audio-tapes. (d)

Throughout the process, I took notes as a back-up to all audio recordings during think-aloud protocols. I then repeated the same steps (a-d) with the Course 10 reader. (e) Finally, if there were any tweeners in the session—students pushed down by the Course 11 reader, but up by the Course 10 reader—I gathered both readers together and recorded a conversation between them as they shared their initial thoughts and decided on the best possible placement for each tweener.

Following each reading session. As promptly after each session as possible, I transcribed the think-aloud protocols for each session and submitted transcripts to each placement reader for member checking (Lincoln & Guba, 1985, p. 314) where they could review what they said, alter it to better capture their meaning, or approve it. Once all original placement sessions were completed and the audio tapes were transcribed and approved, I removed non-participating students from the transcript, and I destroyed the audio recordings. The initial plan for data collection during this phase included steps for follow-up interviews with readers, if necessary, to ask follow-up questions or clarify criteria that emerged during analysis, but no follow-up interviews were conducted with placement readers.

Phase Two: The First-Semester Course

The second phase of data collection began at the start of the Fall 2013 semester courses. This phase of data collection focused on securing the participation of as many of the eight Course 10 and Course 11 teachers as possible. I hoped that at least 50% of the teachers agreed to participate, although higher numbers would give a fuller picture of different teachers' course criteria. Ideally, 75% to 100% participation was preferred. In addition, I was seeking to secure consent from as many students as possible, but I hoped

to have at least 50 students in total (20 in Course 10; 30 in Course 11) with somewhere between 4-6 participating students in each teacher's class. Since students could withdraw their consent to the study at any time, I hoped to conclude the semester with no less than 30 students in total (12 in Course 10; 18 in Course 11). For this phase of the study, my goal was to seek course teachers and students' course work for study to get a sense of the criteria with which teachers assess incoming students, so I could determine how placement criteria align with course criteria, as well as secure some sense of the teacher's perception of the adequacy of placement and the placement prompt.

As outlined above, the data collected during this phase of the study included an interview with course teachers that included a discussion of the placement prompt since, as discussed earlier in connection with Murphy and Ruth (1993), how readers understand the prompt might differ from each other, students, and the intended meaning. But it also included course teachers' think aloud protocols taken while reading first their students course diagnostic essays and later a formal piece of writing in the course, a survey on placement adequacy, copies of all student work discussed, and copies of course documents (syllabus, major assignments, grading rubrics, and diagnostic essay prompt). In what follows, I outline the course sections available for study in this research project and decision making on eliminating certain special sections. Then, I describe how informed consent was secured from course teachers and students, and the number of participants consenting to the study. Finally, I describe the data collection steps for a total of three interactions with teachers, as well as steps taken after each interaction.

Course sections available for study. The schedule for the Fall 2013 semester included seven sections of Course 10 with five different instructors (including one section

taught by myself). This left six sections of Course 10 and four different instructors as possible participants. For Course 11, the schedule included eight sections of the course with five different instructors, one of which who was also a Course 10 instructor. This meant there were 14 total possible sections and eight possible course teacher-participants in total from both classes when the one teacher who taught both courses was counted only once.

In addition, there was a section of Course 11 taught online and an honors section as well, and both of which were removed from the list of possible participants. The decision to remove the online section was simply one of practicality; it would be difficult to meet with the teacher or obtain informed consent from the students in an online environment. The honors section was excluded because it contained only a handful of students and, by its very definition, should include students whose writing attributes would make them atypical of standard Course 11 students, in effect potentially skewing the resulting criteria if included.

Informed consent and number of participants. To start phase two, I initiated contact with each teacher prior to the start of the fall semester by emailing them about the study (see Appendix G). To this email, I attached an informed consent form (see Appendix H) and asked to meet with each teacher to discuss the study and the form. As with the placement consent form, this form explained the study's purpose and value, and it noted that there were no known risks to teachers, but that participation could involve increased reading time since teachers would have to voice their thinking on student's work. While there were minimal discomforts, there were potential benefits; through the study, teachers could come to better understand how they read student work, and this was

explained to the potential participants. Seven of the eight possible course teachers agreed to participate. The Course 10 teachers included Adele and Daria, who each had two sections, and Hannah and Theresa who each had one section. The Course 11 teachers included Janet and Theresa, who each had two sections, and Odysseus and Terence who each had one section.

During my meeting with the teachers, I asked for permission to come to their class during the first or second week of the semester to secure students' consent as well. In participating teachers' classes, there were a total of 205 students registered, 101 in Course 10 and 104 in Course 11. With the teacher's permission, I went to their classes and distributed two informed consent forms to each student, so they could keep one for their records if they chose to participate (see Appendix I). The consent form explained the study, its value, what was being asked of students in phase two, how confidentiality would be handled, the potential risks, and the right of students to consent or not. I discussed the study and the form with the students, answering any questions as necessary. Students were instructed to place their completed or blank form (depending on their choice to consent or not) in the same box upon exit from the class so that no one in the class, including the teacher, would know if they chose to participate or not.

Students who consented were giving permission for me to watch and record their teacher as they read the student's diagnostic essay and first writing assignment; to read, copy, and use this work in the study; to access their final course grade and portfolio at the end of the semester, and, as I described in phase one while discussing the backup plan for securing the minimum number of placement participants, students were agreeing to let me pull their placement essay and placement records and to perhaps even have their

placement essay reassessed. Since the reading of student work was already a part of the course and since the focus of the study was not the students' work individually but the range of criteria under discussion by the teachers early in the course, the procedures of the study presented only a minimal risk to students. This risk was diminished by keeping the identity of participating students confidential from course teachers. Also, the design of the study attempted to minimize risk by employing think-aloud protocols where teachers were encouraged to read as they would naturally as they voiced their thoughts aloud (Ericsson & Simon, 1984). Additionally, it helped as well that the focus in reading the diagnostic and the first writing assignment was not on grading but on the criteria that determine placement and that are important to the work of the course itself.

One-hundred and eighteen students in total agreed to participate in the second phase of the study—59 in Course 10 and 59 in Course 11—a number that more than doubled the minimum goal of 50 participants and represented 48.76% (118/242) of the total number of students enrolled in first-semester courses that semester (50.43% of enrolled Course 10 students and 47.2% of enrolled Course 11 students). The single student who consented to participate initially at placement was not part of any of the participating teachers' courses, and thus was not part of this phase of the study. Additionally, this group was made up of 52.54% female students and 47.46% male students with 17.80% of the total being international students. The make-up of the group was not that dissimilar from the population of first-time, full time freshmen overall, having about 4% more male students and consequently 4% fewer female students than the population of first-time, full time students. The international student numbers were nearly exact with less than 1% difference in overall makeup.

First teacher interaction—interview. The criteria under discussion during the courses were ascertained in a number of ways all centering around three teacher interactions. The first interaction was completed during the second to fourth weeks of class. I conducted and audio-recorded an interview with each teacher to learn about the teacher's background and experience; how he/she understood the program, course, and its principles; how he/she understood the placement prompt and its relationship to the course; and how he/she had designed the assignment sequence to teach course principles. Additionally, the teacher chose a pseudonym during this interview. (see attached interview questions in Appendix J). While these questions were prepared ahead of time, I kept the interview semi-structured and the questions open-ended (Denzin & Lincoln, 2008) to allow the reader to share information that was relevant to him or her that I did not anticipate.

Second teacher interaction—think-aloud session on diagnostic essay. Also during the second to fourth weeks of the semester, I met with teachers in a second interaction. Here, I conducted and audio-recorded think-aloud protocol training (see Appendix K); collected and audio-record think-aloud sessions as the teacher read the diagnostic essays of students; and assigned code numbers to and logged each diagnostic essay as it was discussed. Throughout think-aloud sessions, I allowed teachers to notice what they normally would and describe what they saw as they determined if the student was adequately placed. The only prompting was in response to sustained silences reminding the reader to think-aloud. In addition to helping determine course criteria, the think-aloud protocols were also a way to check the teacher's sense of the adequacy of placement given one impromptu essay. The goal was to learn about the teacher, the

course, and the criteria this teacher used to check placement. In addition, as the diagnostics were read, I asked teachers to complete a survey on the adequacy of placement overall where they indicated which students they felt were prototypical of the course, slightly above, slightly below, should be moved out of the course to a higher course, or should be moved out of the course to a lower course (see Appendix L for a copy of the survey).

Third teacher interaction—think-aloud session on course paper. The third interaction with teachers was scheduled during the fifth to eighth weeks of the semester. The initial planning for this interaction had it scheduled within weeks four to six, since Smith (1993) indicated that getting too close to the midpoint of the semester might align teachers' observations less with placement and more with final course grades and outcomes, but it was not possible to schedule any teachers' meetings during week four since many teachers were not collecting a piece of formal writing that could be read until that week. Other teachers had scheduling difficulties that set up the need for multiple, shorter meetings over the course of a week or two to complete all the think aloud protocols on students' papers. Despite these difficulties, I was able to schedule five of the seven teachers' final interaction inside the six week cut off point, but I had to settle for the final two teachers running past the six-week mark but staying inside the eight-week midpoint of the semester. In fact, only one teacher scheduled any session during week eight, and that was for the fourth and final meeting to finish up the last few papers being discussed from week six of the class.

The final teacher interaction involved think-aloud sessions as teachers again went through think-aloud protocol training and read student work collected near or around

weeks four to six (which was assigned code numbers and logged just as earlier work had been). The goal was to determine from the teacher the criteria that are discussed as teachers develop a better sense of their students. This data was also interesting to test out the student's progress so far given teacher's earlier observations of their diagnostic essays.

In the initial study design for this stage, focused interviews, if necessary, were planned during this interaction as well, if questions arose through coding and analysis of the earlier interactions, but no focused interviews were necessary. In addition, the initial plan for the study included a second survey of placement adequacy using the students formal paper, but given the scheduling difficulty that already pushed this interaction back several weeks and cramped the schedule of teacher participants, I forwent the additional survey since I already had the previous survey on students' diagnostics to give me a number of students adequately placed, which is usually the only time that measure is collected in a given semester anyway.

After each teacher interaction. Following each interaction, I completed record keeping, transcription, and member-checking activities similar to phase one. Thus, I took notes during all interviews and think-aloud sessions as back-up to audio-recording; copied, numbered, and logged each essay, piece of work, or prompt under discussion during think-aloud sessions for future analysis, and so it would be clear what order they were discussed in on audio-tapes; transcribed each interview and think-aloud session promptly afterwards; submitted transcripts to each teacher so they could review what they said, alter it to better capture their meaning, or approve it as per the outline of member checking (Lincoln & Guba, 1985, p. 314); Once all sessions were finished, transcribed,

and member-checked, I destroyed the audio recordings. The initial plan for data collection during this phase included steps for follow-up interviews with teachers, if necessary, to ask follow-up questions or clarify criteria that emerged during analysis, but no follow-up interviews were necessary with course teachers.

Phase Three: Reassessing Placement and the End of the Course

The third phase of data collection took place during weeks twelve to sixteen of the Fall 2013 semester. This phase focused on two tasks—having repulled placement essays reassessed by the placement readers and completing the collection of the final pieces of data related to the end of the first-semester course. The data collected during this phase of the study included, on the placement side, placement readers' think aloud protocols taken while reassessing students' placement essays and copies of the student essays themselves. The end of course data included copies of students' final portfolios and an end of term sheet filled out by the teachers, which reported the students' final course grades and outcome.

In what follows, I provide the number of participants for this phase of study, indicating that consent was secured during previous phases of the research study. Then, I describe the process for repulling student placement essays. Next, I detail the procedure for redo placement sessions, including procedures for preparing repulled essays for the redo placement reading sessions, procedures during the sessions, and procedures following the sessions. Finally, I describe how the final pieces of information on the end of the course were collected, including copies of student work, grades, and final course decisions and reports.

Informed consent and number of participants. Students consented to the tasks associated with this phase in previous phases of study. One-hundred-and-nineteen students, a total from both phase one and phase two, consented to having their placement essays read, and 118 from phase two consented to allow their final course results to be given to me by their teacher.

Pulling students' placement essays for redo placement sessions. My first focus during this phase was on repulling students' placements and planning redo placement sessions. As I explained earlier, if I did not obtain an adequate number of participating students during the placement sessions of phase one (at least 50 essays fairly distributed in Course 10 and Course 11), I would repull the placement essays and any recordings and notes from placement readers on the essays of phase-three participating students, and I would schedule additional meetings with placement readers to conduct think-aloud protocols as they blindly read these essays again. In this way, I would make sure that the criteria I uncover at placement was based on a sufficient number of placement essays such that the essays were not outliers but representative of the larger group of essays.

One-hundred-and-nineteen students—both those who agreed to participate during both phase one and phase two—consented for me to collect their placement records and to watch placement readers read and assess their placement essay again, as needed. Thus, I provided the office in charge of placement testing with a list of their names so the office could copy students' placement essays for me. However, when the office in charge of placement testing pulled copies of the essays from student files, 11 students' folders could not be found or their folders contained no placement essay. The office clerk noted on the original list of names I provided her which students were missing copies along

with a reason, if available. There are a few scenarios that could cause a student to lack a file or a placement essay. First, as I've noted previously, ESL students write their placement essays in their final ESL writing course, and copies are not always forwarded to the office. ESL students account for six of the 11 missing essays. Second, students in a dual enrollment program at Cardinal College can take Course 11 for college credit while also enrolled in high school, but they are not required to write a placement essay since they are granted admission to the dual enrollment program through the recommendation of their high school teacher and high school transcript review. Dual enrollment students account for two of the 11 missing essays.

Of the remaining three missing students, one student had no placement essay and no notation next to her name in the list at all; it appeared that the student had been mistakenly skipped in the folder search. The final two students had a notation next to their names that indicated their folders were missing from the files altogether. One student was a transfer student and his file was inaccessible because it was likely being held in the registrar's office while his transfer credits were being verified and uploaded into the official electronic student record. One student was a returning student, and his file was likely being held with whatever advisor or department he was associated with when he was last enrolled. The result was 108 students' placement essays available for the redo placement think-aloud sessions.

Redo placement session procedures prior to reader arrival. Prior to scheduling and holding redo placement think-aloud session, the 108 repulled students' placement essays were logged and specific identifying information and details were recorded, such as the original placement decision made about them, if they were tweeners

in the original placement session, the original session they sat for, the teacher and section they were enrolled in, if they were ESL students, and if there were any notes on their placement indicating they would have been placed in Course 11A in prior years. Next, the essays were copied with identifying marks and any decision notations removed so they could be read blind—without readers knowing either their initial decision or the identity of the student. As the identifying marks were removed, I created an identification number for every essay, recording the number on the top of the essay as well as in a log for notes.

Next, I distributed the essays into three redo placement sessions, accounting for the earlier noted characteristics. I chose three sessions partly because of the difficulty of scheduling more sessions than three at this stage in the term, but more importantly because three sessions with roughly 36 essays per session would not be an unusual amount for any given actual summer placement session. While the last few sessions over the summer were not as heavily attended, there were three out of seven original placement sessions with over 40 essays collected and one additional session out of the original seven that had 33 essays. Students were first distributed into these three sessions so that no one redo session contained a large collection of the placement readers own students or students from one particular section of the reader's class in case these students handwriting or writing characteristics would cause the reader to identify them as their own student.

Adjustments were made in the distribution of students in redo sessions to evenly distribute ESL students since the majority of ESL students' essays are read in the session following the completion of the spring ESL courses and only a small number of ESL

students typically sit for a given regular placement session over the summer. Additional adjustments kept the grouping of students as close to actual original placement sessions as possible in terms of the percentages of original placement decisions, of tweeners, and of students designated old Course 11A students. Original placement sessions varied in their exact percentage of Course 10 students, ranging from 41.67% at the lowest to 54.55% at the highest for any one session with exactly 50% Course 10 distribution overall, once ESL student sessions were factored in with summer placement sessions. The closest I could come to staying inside the range presented by the summer placement sessions meant Redo Session 1 would have 47.22% original Course 10 essays, Redo Session 2 would have 54.29% and Redo Session 3 would have 54.05%.

Procedure during redo placement sessions. I followed the same basic steps for the three redo placement session as I did for original placement reading sessions. (a) I offered the opportunity for think-aloud protocol training with the Course 11 reader (see Appendix F) at the start of the session. (b) I collected and audio-recorded think-aloud sessions as the Course 11 reader read all student essays and made decisions about acceptance or rejection from Course 11. I allowed the reader to notice what she would and describe what she saw as she normally would. The only prompting was in response to sustained silences, reminding the reader to think-aloud. (c) During the think-aloud sessions, I logged each essay and placement decision so it was clear what order essays were discussed in on audio-tapes. (d) Throughout the process, I took notes as a back-up to all audio recordings during think-aloud protocols. I then repeated the same steps (a-d) with the Course 10 reader. (e) Finally, if there were any tweeners in the session—students pushed down by the Course 11 reader, but up by the Course 10 reader—I gathered both

readers together and recorded a conversation between them as they shared their initial thoughts and decided on the best possible placement for each tweener.

Following each redo placement session. As promptly after each session as possible, I transcribed the think-aloud protocols for each session and submitted transcripts to each placement reader for member checking (Lincoln & Guba, 1985, p. 314) where they could review what they said, alter it to better capture their meaning, or approve it. Once all original placement sessions were completed and the audio tapes were transcribed and approved, I destroyed the audio recordings.

Collecting final course data. The second focus in this phase was on collecting final results of the semester in terms of course grades, outcomes, and reports. At the end of the term, the department conducts an end of term review (EOTR) session on the Friday of finals week where, among other things, teachers bring their students' portfolios to be reviewed by other teachers in the program and pass/fail decisions, as well as decisions on Course 10 students permitted to jump to Course 12, are made collectively to ensure a consistency to the standards of the program. At this meeting, I made contact with each participating teacher and distributed an EOTR course result sheet (see Appendix M) for them to use during the EOTR meeting to record any outcomes of pass/fail decisions or Course 10 jumps. In addition, there was space provided on the sheet for the teacher to record their students' final course grades and a final note of assessment on the students' work. To ensure that participating students' identities were not revealed to the teachers, I had to ask teachers to record all students' final grades on the form. However, immediately upon receiving the form, the sheet was typed with all participating students' names replaced with their identifying code in the study, and all non-participating

students' names removed completely. The original sheets were then destroyed. I asked teachers to return the EOTR course result sheet to me as soon as possible after they finished calculating final grades, which were due the following week.

In addition to distributing the form, I also made arrangements with teachers regarding when I would collect students' portfolios to make copies of participating student work. Dates were set around teachers' holiday vacation plans, with some arrangements only being possible the following month, closer to when classes would resume for the spring. I also made arrangements with the program director to acquire the final semester report which contained the official total pass/fail numbers for Course 10 and Course 11, as well as the total number of Course 10 students who were granted permission to skip Course 11 and go directly to Course 12.

Confidentiality During Data Collection Overall

As discussed throughout, confidentiality of data was maintained by using codes for students and pseudonyms for teachers. Placement readers and course teachers did not know which students were participating or not. Only my dissertation advisor and I had access to master records with subject identifiers. Once all data was aggregated, master records with identifiers were destroyed, as were all audio recordings once they had been transcribed and approved. All other data was and will be retained for at least three years in compliance with federal regulations.

Debriefing Participants

Sales and Folkman (2000) explain that debriefing is a "two-way educational process between the researcher and the research participant" that "provides participants with information about the nature of the research and available results," yet it can also be

used “as a way of managing harm” and misunderstandings (p. 65). Directly following each teacher interaction, as I already said, I followed the practice of member checking (Lincoln & Guba, 1985, p. 314) where teachers reviewed transcripts of interviews and think-aloud protocols to refine their intended meaning. This, in and of itself, was a kind of debriefing as it presented a collection of the participants’ thoughts at once for conscious reflection. But additionally, once the analysis of the data was concluded, I discussed and shared the results with the writing program director and department members to debrief participants and share gained understandings.

Steps of Data Analysis

Now that I have detailed the steps of data collection, I move to outline the steps of data analysis for this study. First, I describe the creation of a master list of identifying codes, information, and relevant labels for each participating student. I describe the creation of this list first since I began the master list as I started collecting data—using my logs and notes from placement sessions and teacher interactions—and the list was maintained, checked, and added to across all subsequent data analysis. Second, I detail the analysis procedures for each category of data—writing program placement records, placement-reader and teacher think-aloud protocols, placement and course documents, teacher interviews, placement adequacy surveys, end of course result sheets, and the writing program end of term report—moving through initial organization and first impressions to later examinations looking for patterns and themes within each data category. I organize here by category of data because each category of data required different procedures specific to its data type and the purpose it would serve in answering

my research questions. Third, and finally, I discuss analysis procedures for comparing and connecting themes, patterns, and information across the categories of data.

Master List

As there were many points of data collection in the first weeks of the study—initiating contact with placement readers and placement takers, collecting placement think-alouds, initiating contact with teachers, and the first interactions of the course—there wasn't time for much in the way of analysis at the beginning. However, one of the first tasks I undertook early on was the creation of a master list of all participants where I could record their identifying pseudonyms (for teachers) or code numbers (for students) and where I could keep track of the data I had collected with important demographic and study-related information about each participant. I made the master list as a spreadsheet, which included all 119 participating students by their assigned code number in the study. The official code number for each student was designed to contain their course and section number in it, so that it would always be clear later which course and section they were in. For example, 1011-02-03 signaled the student was placed into Course 11 (1011), section two (02) and was the third student to consent from that section (03). The one student who did not go on to participate in phase two, and who I had no course information on outside of his placement into Course 11, was coded as 1011-00-01, designating that he had no assigned section number.

After the initial creation, I used the list as a way to keep track of other important identifying information about students as well. For instance, as I was conducting placement sessions, students' essays were given code numbers to signal their order on tapes (for the original placement sessions) or to hide their identity from readers (for the

redo placement sessions). Keeping track of their code inside different think-aloud sessions was necessary because during the original placement, students were not yet assigned an official study code, and during the redo placement sessions, I did not want to use the official study code, which would give away their original decision.

As the study went forward and I reviewed my notes and, later, collected data like the official writing program reports and notes from placement, transcripts, surveys, etc., I added important decisions and labels that were assigned to the student at different stages. For example, if the student's essay was a tweener during a placement session or if a course teacher indicated the student was slightly high in the class, these labels were included in the master list so I could easily find all students who fit a certain category or description if necessary. The final list contained a record of each student's study code and the following items (if those items were collected for that particular student) in the following categories:

1. Original placement information, including the student's placement decision, session number, code in the session, and placement labels (TW for tweener and 11A for Course 11A designation);
2. Redo placement information, including the student's redo session decision, session number, code in the session, and labels, (TW for tweener, 11A for Course 11A designation, and DD for students whose redo decision differed from their original placement decision);
3. Course information and results: the student's teacher, their codes in both of the teacher think-aloud sessions (including if they failed to submit work in time to be discussed in any of these sessions), the survey label their

teacher assigned them (prototypical, slightly below, or slightly above), their submission of a final portfolio, their final course grade, and if they were a Course 10 student who was permitted to jump directly to Course 12 (noted as a jump).

As I discuss the analysis of the data in this study in Chapter Four, I will naturally mention as part of my discussion pieces of the information recorded on this master list in regards to the total 119 participating students as they become relevant. However, referring to students, as I will, by their study code numbers, their demographic information may not always be clear. See Appendix N for a list of demographic information for all 119 participating students, including designations for gender, English-as-a-Second-Language students, and concurrently-enrolled high school students.

Writing Program Placement Records

Having only collected data directly from the final three of seven placement sessions over the summer, it was necessary to collect the writing program's official placement records for all other sessions and its final placement report so I could have a sense of the overall results of placement for the total population of freshman entering the first-semester writing courses that year, as well as specific decisions and labels given to those students who agreed to participate in my study at phase two, but who did sit for one of the final three placement sessions on which I had directly collected data. As I already indicated above, I added the specifics gleaned from these records to my master list of students so I would have a record of students' decision, session number, code number, and labels.

Additionally though, I used the notes and reports to calculate and check the statistics overall. I calculated the total percentage of students placed into Course 10, those placed into Course 11, those tweened, and percentages for two other occurrences. First, a placement reader will, on occasion, indicate on a particular essay more than just the decision they made: they may write a specific teacher's name down as a suggestion for the advising office when selecting sections for a particular student's schedule. I calculated how frequently a teacher was recommended. Second, as I explained earlier in the chapter, there once existed a course between Course 10 and Course 11 (Course 11A), and placement readers occasionally note if they feel a student would have been placed in that course in previous years. Making such a distinction, it is thought, helps the advising office make the best schedule for that student as only some teachers taught the old Course 11A while others did not, and old 11A teachers would be more familiar with 11A students' needs. The percentage of students marked as 11A was calculated as well.

Think-Aloud Protocols

As detailed during the data collection description, two sets of think-aloud protocols were collected from both placement readers as they decided placement (original placement sessions and redo placement sessions), and two sets of think-aloud protocols were collected from seven course teachers as they read students' work in their Course 10 or Course 11 sections (diagnostic sessions and paper sessions). Before analysis, the think-aloud sessions were transcribed and approved by the readers or teachers. Further, as I worked with the approved transcripts, I used them to add notations and labels to the master list, as I mentioned earlier. In what follows, I describe the stages of think-aloud

transcript analysis from noting initial impressions to coding to find patterns relevant to my research questions.

Initial impressions. The first step of analysis took place as the placement and course think-aloud sessions—original placement, redo placement, diagnostic, and paper sessions— were transcribed: I kept lists of impressions and criteria that occurred to me as I worked, noting ideas to which I wanted to return. I also consulted the actual copies of students' writing to clarify confusion. For example, when I couldn't tell if a participant was reading from a student's essay or making comments of his or her own, I looked at the copy of the student's work to see if I could solve the problem. Finally, as I was transcribing and making initial notes, I recorded important decisions made about or labels assigned to students as the teacher talked to order to add them to the master list that I was creating.

Excluding paper think-alouds. Before I proceed, it is important to note here that given the large number of students participating in the study—more than twice as many students consented to participate than required—I made the decision, after transcription and initial analysis, that it was best to post-pone analyzing the paper think-aloud sessions and to answer the research questions here without considering that data. Instead, I used only the original placement sessions, redo placement sessions, and diagnostic course sessions. While the inclusion of the paper think-aloud data was important to the initial design of this study, it was not possible to achieve the level of close qualitative analysis required on that many think-aloud protocols given the number of students whose work had been read. As it was possible to answer the research questions without the paper session data, I decided to save the paper think-aloud data for future analysis or

subsequent validation studies, adding in the analysis of the paper think-aloud sessions and the consequent findings to those presented here at a later date and in the consideration of future validation questions, studies, or publications.

Coding. From these first steps, I moved to more formal coding. Given the first research question of this study to investigate the assumption that Cardinal College places using course criteria, the question this data set addresses most directly, I coded the think-aloud sessions for themes and “significant statements,” and grouped them “into meaningful units” (Creswell, 2007, p. 156) of specific reading and writing criteria under discussion for Course 10 and Course 11. As mentioned earlier, Creswell (2007) suggests that the code emerge from the data, as opposed to using a preexisting code, because then the code can reflect the views of participants inside the research more fully. Following this principle, the development of the code went through several stages.

Additionally, it was natural for sections of transcripts to be recoded as specific codes were combined with others when the criteria of the participants were understood more clearly and as new codes developed during the analysis of later transcripts and sessions. In addition to the reading and writing criteria, placement transcripts specifically were coded for another specific kind of comment. Since Smith (1992/2009) specifically discusses that expert placement readers may also note behaviors of “the writer as a student” as opposed to just the characteristics of the essay—which again Smith claims is privileged course knowledge that helps expert readers make their placement decisions—I also coded for this type of statement in the placement transcripts (p. 183).

Further, all think-aloud transcripts were coded for a number of other kinds of comments. For instance, the transcripts also included off-topic chatter or interruptions by

colleagues or phone calls, comments made by readers or teachers about the specific prompt or assignment associated with the piece they were reading, and questions to me about the procedures they should follow to pile papers or mark decisions as they read. All sets of think-aloud transcripts also exhibited some form of fit-to-course comments. In the case of placement readers, such fit-to-course comments were made when the reader made their decision about the student (deciding to place them in the course or push them up or down), or when they made their decision while noting that the student did not fit neatly in either course. For course teachers, such comments were made during the diagnostic reading when teachers had to decide if students were prototypical of the course or some degree above or below the course, including if they were improperly placed. These types of comments were coded for as well.

Major codes. From the above discussed coding processes, 20 themes emerged. 14 of these themes directly correlated to criteria of reading, writing, or thinking that readers were discussing with student essays. Six of the themes were not criteria themselves, but instead, they represented other kinds of comments made in the transcripts. Table 1 describes all 20 themes.

Additional notations. Now that I have described the coding themes, I should detail an additional layer of analysis which integrated notations on three other important observations about the think-aloud comments. First, as I coded, subcategories of the coding themes (subcategories of the criteria really) emerged and were noted. When readers sought to describe or name the issue a student had with a particular criterion (most notably with codes 9 and 10 in Table 1, reading issues/quality and assignment task), or when patterns were discovered in the kinds of comments given (most notably

Table 1

Think-Aloud Coding Themes

Number	Name of Theme	Description of Theme
1	Language	Writers' word choice, vocabulary, & syntax
2	Grammar	Writers' grammar, punctuation, & spelling
3	Academic rhetorical conventions	Writers' naming the author, naming the texts, & audience awareness
4	Thesis/project	Writers' viewpoint, position or argument overall
5	Framing/moves	Writers' moves, framing, or paragraph/section purpose
6	Essay/ paragraphing knowledge	Writers' understanding of essay structure, paragraph structure, & paragraph breaks
7	Length/amount of writing	Overall length of the response as well as amount of quote versus writers' content/words
8	Working with text	How the writer works with quotes—including using, choosing, marking, setting up, & explaining quotes
9	Reading issues/quality	Writers' understanding of the texts
10	Assignment task	Writers' understanding or performance of the task called for by the assignment
11	Use of required texts	Writers' use of the required materials called for by the assignment
12	Print conventions	Writers' handwriting, indentation, & margins
13	Writing process	Writers' knowledge of the writing process
14	Thinking	Writers' ability to think about, extend, or connect with ideas from the reading or on the topic
15	Course privileged knowledge	Placement reader comments showing knowledge of student behavior in course (Smith, 1992/2009)
16	Fit to course	Placement reader makes placement decision and/or comments that a student is between courses or shows characteristics of both courses
		Course teacher ranks student as prototypical or some degree of above or below the course
17	Prompt	Reader/Teacher comments on prompt or testing constraints
18	Questions	Reader/Teacher asks a question on think-aloud processes, marking papers, or organizing piles as reading
19	Clarifications	Researcher clarifying procedure or asking reader/teacher to clarify meaning
20	Off topic	Reader/teacher chatting, tangents, & direct addresses to the researcher, as well as outside interruptions

with code 15, course privileged knowledge), subcategories were noted. For example, while the majority of discussions on reading quality named the issue broadly as “misreading,” on occasion the particular way of “misreading” was explained further. For example, in Redo Session 1, Theresa described student 1010-05-04’s misreading as “making things up” to emphasize the way the student seemed to follow a line of thought not in the text, although seemingly related, but attributed to the text. The subcategories of criteria in the data will be discussed in more detail in Chapter Four in relation to particular findings.

Second, readers’/teachers’ discussion of a particular theme or criterion would fall on a continuum of ability with the comment often having a negative or positive connotation, which was noted as well. For example, a reader could discuss reading quality negatively, noting misreading, as Adele did on student 1010-01-07 in Redo Session 1, saying “Okay, this is a misreading of the text here.” Or, a reader could discuss reading quality in positive terms, noting an important idea understood, as Theresa did on student 1011-03-09 in Redo Session 1, saying the student “understands that Graff is using an analogy” and that the student “gets what Graff is saying about...things being hidden.”

Third, some readers’/teachers’ comments occasionally gave me the impression that they were having difficulty making a placement or fit-to-course decision or were reluctant to suggest a student be moved from their course despite other comments that suggested they felt the student was a poor fit. As I coded the transcripts, I flagged these instances so I could count how frequently they occurred and keep track of these students.

Tallying codes. Once coding was complete, I began working with the transcripts to tally how frequently certain criteria were discussed to determine their relative

importance to specific courses or student groups. Rather than tally the total number of times a particular criterion was discussed overall by counting each mention of a criterion, even if the same criterion was discussed several times on one student, I decided to count the total number of students about which a particular criterion was discussed. I made this decision for several reasons, first of which was its ease. As teachers discussed particular students, it was difficult to distinguish what constituted one versus more than one comment on a particular criterion if, as was often the case, the teacher discussed a reading or writing issue at length, describing what they observed over time. Marking where one comment started and another ended was problematic in these cases.

Second, and more importantly, counting the criteria by number of students gave me the statistic I was more interested in knowing—for how many students was a particular criterion a factor—without having to do additional mathematical calculations to make the tally usable. One other aspect of the coding was tallied as well. Since teachers' comments often took on a positive or negative connotation (or a mix of positive and negative on specific criteria), I not only counted the number of students about which a particular criterion was discussed, but I counted if that discussion was positive, negative, or mixed overall on each criterion.

Distinguishing problematic placements from unproblematic placements. As I looked over the transcripts, noting interesting moments and considering the tally of criteria I had built, I began to think about Smith's (1992/2009, 1993) assertion that it was natural for some students to fall between existing courses as they had been designed, and that even when read by expert readers, these students would consistently pose problems for easy and neat placement decisions. Tweeners, as Smith (1992/2009) theorized, were

not a clear fit into the outlines of existing courses. Further, to Smith, what would typically be seen as disagreement about where to place these students could actually be read as agreement that they didn't fit clearly into an existing course. Since I had built a master list that tracked students' problematic-placement designations (tweener, Course 11A, or different decision), I wanted to account for those designations as I examined the placement think-aloud transcripts for placement criteria. And, since the master list also tracked fit-to-course designations made by course teachers, where they indicated if students were prototypical of the course or some degree above or below the course, I wanted to account for those designations as I examined the course think-aloud transcripts for course criteria. Therefore, I had to use these designations to reorganize the think-aloud data.

Before I detail how I reorganized the think-aloud data, let me explain a bit more about why I felt the criteria of placement into each course and the criteria of each course had to be analyzed this way. One of the primary goals of this study was to uncover and compare the placement and course criteria to determine if placement readers made course-related decisions. Students who, for example, were cleanly and consistently placed would make good representatives for determining the placement criteria for their corresponding courses, as they had been placed into their course in both of two separate sessions without any noted problems. Alternately, students whose placement showed problems or who were not consistently placed would likely be poor representatives of a particular course's placement criteria since, based on Smith's (1992/2009) theory, they would exhibit a mix of criteria or some other set of criteria entirely.

Therefore, analyzing the data with problematic and unproblematic students separated, in theory, would give me the separate data sets with which to answer my research questions since the inclusion of students who were not a clear fit for the courses would likely distort the criteria of the courses. Additionally though, by separating the data in this way, I could directly test Smith's theory with my own set of data since I could look for differences in criteria between students who were easily placed and those who were not. The problematic students were not ignored though. Their data was still analyzed to uncover the criteria related to the boundaries of and space between the two courses, an additional and valuable understanding for Cardinal College's placement process.

For the reorganization of the placement think-aloud data, I followed the following three steps: First, I divided students into the two previously mentioned groups: unproblematic students and problematic students. As I've suggested, unproblematic students were those easily placed in a course with no difficulties while problematic students were those whose master list record indicated that the student presented placement-decision difficulty. To be specific, placement-decision difficulties included students who were tweeners in a session, students who were labeled as old Course 11A students (indicating they showed signs of the old course that was between the two existing courses), and students who received a different placement decision in the redo session than their actual course placement.

Second, I divided each of these lists into two, separating original session designations from redo session designations and making the two lists into four lists: unproblematic original session participants, problematic original session participants,

unproblematic redo session participants, and problematic redo session participants. Since the redo placement sessions were artificial, in that no actual placement decisions were being made and the readers knew that, I wanted to make sure I could distinguish a student who was a tweener in an original session from one in a redo session in case differences emerged in these two groups. Third, I divided each of these four lists by two based on the students' official (and original) placement decision into Course 10 or Course 11, which resulted in eight lists:

1. Unproblematic original session participants placed into Course 10;
2. Problematic original session participants placed into Course 10;
3. Unproblematic original session participants placed into Course 11;
4. Problematic original session participants placed into Course 11;
5. Unproblematic redo session participants placed into Course 10;
6. Problematic redo session participants placed into Course 10;
7. Unproblematic redo session participants placed into Course 11;
8. Problematic redo session participants placed into Course 11.

Using these eight lists, I could analyze the separate course criteria while distinguishing problematic students from the remainder of students, and acknowledging determinations made originally from those in the redo session.

The reorganization of the course think-aloud data was a little different. While I wanted to note which students were designated as above or below a particular course, I did not want to assume that all teachers of a particular course would make these determinations based on the same criteria. Therefore, I separated students first by teacher, and then, under each teacher's list, I made three separate lists: students described as

prototypical, those described as slightly above, and those described as slightly below.³

Once these lists were made, I grouped all Course 10 teachers under one list and all Course 11 teachers under another, so that once individual teachers' criteria were determined, shared course criteria could be uncovered. Just as I did with the placement transcripts, I first analyzed the course students described as prototypical separately from those described as slightly above or below so I could build a set of criteria for the course that was not affected by students the teachers themselves identified as atypical.

Discerning patterns within the data set. With the reorganization of the data and tally completed, I began analyzing what readers and teachers said, looking for patterns to determine the placement criteria for each course and then to determine the course criteria. Looking at all the students on each of the eight placement lists in turn, I looked for patterns in those students' criteria. For example, what criteria were shared by all, most, or many of the unproblematic Course 10 students in the original session? How about Course 11 students in that session? Once I had a sense of each grouping, I looked for shared patterns between both sessions.

I want to stress again, that I used the unproblematic students to outline the placement criteria for Course 10 and the placement criteria for Course 11, but the problematic students, from both courses together, were analyzed to determine the placement criteria that represent the edge and/or space between these courses. Next, once I had a sense of the separate placement criteria of each course, I reexamined the

³ It is important to note here that there were two other possible designations for Course 11 students—students who should be moved to the course above, the honors section of Course 11, or students who should be moved below to Course 10—while Course 10 students had only one other option for students who should be moved to Course 11. However, none of the participating students in the study were flagged by teachers as needing to be moved to a different course, so those descriptions were irrelevant to the students participating in this study.

transcripts for relationships among the criteria, for example, if specific criteria co-occurred in interesting ways. Finally, I analyzed the data, looking for indications of specific developmental relationships between the criteria of Course 10 and that of Course 11.

The process of discerning patterns for the diagnostic transcripts was not all that dissimilar from the placement process. The ultimate goal was to understand the criteria of the course teacher. To build to that understanding though, I looked for patterns shared across students in one teacher's courses in each category of students—slightly above, prototypical, and slightly below—before combining all teachers of a certain course together to look for shared criteria of the course. I was careful to check the patterns in individual teacher's discussions of students first so that I could see what students shared, but also so I would be aware if any one teacher of a course seemed different than his or her counterparts. While academic freedom permits and strong teaching includes different pathways to meeting teaching and learning goals, teachers working in a program, to be effective, have to have some shared vision and understanding of their work if they expect students to take the knowledge and experience gained in early courses into later courses. Finally, just as I did with the placement criteria, I reexamined the course transcripts for relationships among the criteria and for indications of specific developmental relationships between the criteria of the two courses.

Once I had completed the analysis of the placement and course think-aloud transcripts separately, I was able to compare the two sets of criteria to answer my research question directly. Specifically, I put the criteria for placing a student into Course 10 against the teachers' Course 10 criteria to discern similarities or relationships between

these sets of criteria. And, I did the same thing with the placement and course criteria for Course 11, as well as the placement and course criteria for students on the edge of the courses. After directly comparing the criteria I had uncovered, I examined sets of students in-depth. For instance, there were a set of 26 students for which I had both original and redo think-aloud data and a set of 21 students for which I had original placement, redo placement, and diagnostic think-aloud data. Looking at these sets, I could examine the issues discussed with particular students in each transcript as another way to compare the placement criteria and decision to the course criteria. More will be said about these comparisons between placement and course criteria as I discuss the results specifically in Chapter Four.

Placement and Course Documents

As a part of data collection, I already described how the placement prompt and readings, as well as each teacher's course documents (syllabi, grading guides, diagnostic assignments and readings, and formal paper assignments), were collected. This data was analyzed in two different ways. For my research question, which asked specifically about the relationship of placement criteria to course criteria, the documents were analyzed to uncover explicit and implicit goals for the courses, including specific reading and writing behaviors students should learn or demonstrate. By analyzing the documents in this way, they could offer another set of data that specified reading and writing criteria for the courses. This information was compared to the criteria from the think-aloud transcript analysis, as a way to use multiple data points to triangulate or check for consistency. The documents were analyzed in other ways that pertained to my interest in getting preliminary data that may help with the design of a future study on the placement prompt,

looking for common tasks or conceptions of reading and writing in course documents to compare to placement.

Interviews

As part of the data collected, each participating teacher was interviewed about such things as their background, their understanding of the program and course goals, and their approach to their class at the start of the semester. Each interview was then transcribed. To analyze the interviews, I first separated teachers by course. Then, I pulled out each teacher's description of his or her corresponding course. I wanted to use these responses to get a sense of the curriculum, so I identified common themes and patterns for each course. Next, I pulled out each teacher's discussion of specific reading and writing expectations they had of students at various stages of the course and looked for themes or patterns again. Just as with the course documents above, I triangulated what teachers said about their expectations here to the findings from the think-aloud protocols and course documents, checking to see if the criteria of reading and writing here were consistent to those from other data sets.

Survey

As explained earlier, teachers would check placement by administering a diagnostic essay in the first week of the class. Since placement decisions at Cardinal College are made on the basis of one on-demand essay, the program collects this second sample to make sure the decision made at placement still holds given a new sample on separate texts. As teachers completed their think-aloud sessions while reading their students' diagnostic essays, they also completed a survey on the adequacy of placement, which counted how many students they felt were prototypical of the course, how many

were slightly above or below, and how many they felt should have been placed in the course above or below their own (if applicable). The first use of this survey information was in the development of the master list of participants. Students flagged as above, below, or slightly above or below had these course designations added to their master list record.

The primary purpose of collecting this information though was in the calculation of a quantitative measure of placement adequacy—for triangulation with other data in the study. To calculate this measure, I combined all of the teacher’s survey results, including the number of students who were not reassessed because they failed to submit a diagnostic essay, to have a total count of students in each category. Then, I calculated the official numbers on the adequacy of placement for the semester. Specifically, I calculated the percentage of students overall that were assessed as properly placed, the percentage improperly placed, and the percentage not retested. For comparison sake, since Smith (1993) said that teachers are often reluctant to move students even if they feel those students are misplaced, I calculated the percentage of students properly placed and the percentage improperly placed in two different ways. In the first, I counted the students designated as “slightly” off the course—either above or below—in the properly placed category. Then, I recalculated the numbers a second time with these students in the improperly placed category. The official number for placement adequacy was the first calculation where students showing slight designations were still properly placed as there is always some natural and acceptable variation in students’ fit to a course. I wanted the second count though as a comparison number on the worst case scenario I could imagine where all students that teachers flagged as slightly off were actually misplaced.

End of Course Result Sheets

At the end of the semester, there were several pieces of information reported on the course result sheet collected from teachers. Students' final grades were recorded on the sheets, as were notes and final decisions from the End of Term Review (EOTR) meeting on Course 10 students who were being recommended to skip to Course 12 and on any student whose work was on the border between passing and failing. Also, the sheet included teachers' final notes on each student's final portfolio.

Grades were analyzed as one separate and distinct piece of the data collected here. I made a count of final grades for each participating student, adding it to the master list, but also calculated the percentage of participating students who failed, passed, jumped, withdrew. I also used information to tally the number of each letter grade teachers assigned to calculate the median grade (as I did, I separated grades not just by letter, but by plus and minus as well). I calculated the median because, as I explained earlier, Smith (1993) cautioned that teachers may not have the same criteria for determining specific letter grades, particularly given how much student behavior is factored in to the final grade. Therefore, Smith suggested knowing the median grade so students can be sorted as above, below, or on this median grade.

In addition to just grade calculations, I worked specifically with the data across these sheets—grades, comments, and decisions—to add notes to the master list. I indicated specifically which participating students jumped from Course 10 to Course 12; which withdrew, passed, or failed; and which failed specifically because they did not submit a final portfolio. Then, using the teacher comments, the notes on EOTR conversations, and course decisions and grades—i.e., looking across the pieces of data

provided on the course result sheets as a single source of data—I looked for patterns that might help me determine reasons for specific course decisions. For instance, by looking at the teacher’s comment on Course 10 students who were permitted to go directly from Course 10 to Course 12, I checked for patterns in the reading and writing criteria noted in these final comments. As necessary, I reused the think-aloud protocol codes of reading and writing criteria developed earlier for this analysis of teachers’ final comments on student work (see Table 1 provided earlier).

Writing Program End of Course Report

While the final course result sheets provided me with final grades for participating teachers and students, I had to collect the final end of term course report from the director of the writing program to have the final grades for all teachers and students in the program. Using the data recorded as part of the writing program’s reports, I made the same calculations that I did with participating students final grades, only now for all students. I counted the number of students assigned each letter grade designation (noting plus and minus as separate grades), and I calculated the percentage of students who failed, passed, jumped, and withdrew. Finally, I calculated the median grade.

Tracking Across the Data Sets

In addition to analyzing within each single data set, as I’ve detailed in the sections above, I was interested, in tracking across the data sets for interesting and revealing patterns related to my research questions. I used the master list as a starting point for tracking certain student groups across all phases of data collection and data sets. Specifically, I tracked those students who had a placement designation that indicated they did not fit neatly into one course: students flagged as Course 11A by one or more

placement reader, students who were tweened in either their original or redo placement session, and students who had a different decision made about their placement in the redo session than in the original session. I was also interested in tracking those students who had a course designation that indicated they were not a prototypical fit for the course into which they were placed, having been flagged as slightly below or slightly above the prototypical student of the course. And, I was interested in tracking those students who had specific course outcomes of interest: students who withdrew from their course, students who failed their course, or Course 10 students who jumped directly to Course 12.

As I could see all key information on a particular student at one time on the master list, I first checked for co-occurrences of data traits or attributes. Specifically, I checked to see how many students had multiple and non-contradictory designations suggesting a fit-to-course issue, which would signify more strongly that the student did not fit neatly in Cardinal College courses. For example, there were a number of students who were designated as slightly high by their Course 10 teacher who were also tweeners or flagged as Course 11A during placement. This analysis also uncovered students who had only a single designation suggesting a fit-to-course problem and those who had contradictory indications. For example, a placement reader flagged a Course 11 student as old Course 11A, which suggested they might be below or slightly below Course 11, but the teacher reading their diagnostic said they were slightly above Course 11. For all of these students—those with co-occurrences, those with single designations, and those with contradictory designations—the master list helped me notice the issue, and then I

revisited the other data sources looking for what I could learn about particular students or about all students with similar designations.

To offer an example, using the master list to locate all students who were indicated as slightly below their course, I returned to the final grades and looked at outcomes to see if these students struggled or if they were able to handle the course without problems. I also looked at patterns in those outcomes, checking to see if the final grades for slightly below students were lower overall than prototypical students, even if they were able to complete the course.

Further, I inspected the criteria for student groups of interest to the study more closely, examining students who were tweeners, who were designated as old Course 11A, who had opposing decisions made about them in the two placement readings, who were flagged by their teacher as atypical of the course, who failed the course, or who jumped after Course 10. Here again, I was checking for patterns that I could use to understand more fully the relationship of placement criteria to course criteria.

Strengths and Limitations of Data Collection and Analysis

With any study, there are strengths and limitations to the data collection and analysis. Before concluding the chapter, I want to acknowledge the particular strengths and limitations of the data in this study. First, I consider key strengths of the data collection, discussing the number of participants. Then, I discuss several limits to the data collection, particularly related to redo placement sessions, acknowledging how I tried to mitigate these limits before going over issues specific to think-aloud sessions regarding how each participating teacher did with thinking-aloud. Next, I will discuss the limits of the data analysis, noting the difficulties analyzing a large amount of data in-depth and my

own familiarity with the curriculum as an obstacle, but I will then turn to present the strengths of data analysis where I will discuss how I analyzed the data in ways that mitigated the influence of my own viewpoint.

Data Collection Strengths

One of the strengths of the data collection for this study is the large number of participating teachers. Since this study has the goal of understanding the relationship of Cardinal College's placement process to its courses, the study benefits from having such a large representation of the teachers involved in the writing program. 78% of the writing program teachers from the studied semester participated, with only me, as the researcher, and one other teacher not participating in the study. And even though I did not go through the steps of participating as the other teachers did to participate, my views as the researcher are included in the study regardless. Therefore, only one teacher in the writing program from the studied semester is actually absent from this study's data.

In terms of participants, the study also benefits from having a large sample of the studied semester's student population. While the number consenting was not strong for the original placement—only 12% of students placed in this year's placement sessions had their original placement reading studied—the back-up plan allowed for a large number of students' placements to be studied through redo sessions. When redo placement sessions are factored in, 46% of students placed in this year's placement sessions had their placement considered as part of this study. For the course work part of the study, the number is slightly larger: 49% of students in this semester's courses had their course work considered.

Data Collection Limits

One limit of the study, which I mentioned earlier under the data collection procedures, was the limited number of original sessions that were open for study. Since the program was exploring how to have the two readers work together as there was previously only one reader, the writing program asked that the study commence after the procedure was established. Further, given that only a small number of students, 26 out of 224, had their original placement studied, to make up for the deficit in numbers, the study relied on a back-up plan with think-aloud sessions collected from re-run placement sessions for some of its data.

While redo placement sessions solved the problem of participant numbers, it must be acknowledged that redo sessions are not the same as original sessions. The participants, while reading and making the decision with no knowledge of their previous decision, were well aware that this session was not an actual session that would result in placements for the upcoming semester. Therefore, they may have treated it differently than they would have otherwise. Additionally, they may have been more anxious reading in these sessions since they knew a previous decision had already been made, and they would have a personal stake in being consistent. To attempt to limit anxiety and keep the redo sessions as like real placement sessions as possible, they were run with the same procedures, the directions were identical, and they were held in locations where placements had been read before. Further, as I already discussed, I put together redo session groupings where the number and breakdown of students in various courses was as close to actual sessions as possible.

Finally, I want to discuss limits associated directly with the process of thinking aloud. As there were seven different teacher participants in this study, I was able to get a good sense that thinking-aloud comes easier to some than others. Some of the participants struggled with the think-aloud process, perhaps uncomfortable with the process of verbalizing their thinking in general, or perhaps there was an aspect of participant reactivity (Maxwell, 1996) where they felt more guarded under the scrutiny of study. While training and reminders to think-aloud helped participants improve and stay on track, there were some noticeable differences in participants think-aloud tendencies. In what follows, I will discuss the issues I noted with teachers' think-alouds.

Adele and Theresa's think-aloud sessions. I have a very close relationship with Adele and Theresa, as at various times I have worked with both ladies to write curriculum and assignments, design lessons, and theorize on students' learning. Since my relationship with both of these women involves asking for and giving advice and interpretations, as I worked with both Adele and Theresa in think-alouds, I noticed a tendency for them to be very aware of me in the room and occasionally talk to me directly, and I noticed that I struggled at times to redirect them back to thinking-aloud. This was more noticeable with the course sessions than it was with placement sessions, and more pronounced with Theresa than with Adele, as Adele had more experience with think-aloud protocols from her own graduate studies and would self-correct more quickly.

Ericsson and Simon (1984), as I discussed earlier, suggest that the researcher be out of the view of the participant so participants complete the process as they naturally would, which would likely have helped limit this issue, but with Cardinal College's space

limitations, that was not possible. There was no place for me to go where I could be out of sight, yet monitor for recording problems and for taking notes. While comments directed to me as the researcher can be seen as a problem, and while I did my best to ignore or redirect these instances, I want to offer my own sense that there is another interpretation for these moments. Contextually these interruptions may not be a form of interference with the natural process of these participants reading, particularly in regards to their reading process on students' course work, as both these ladies and I frequently read and discuss course work together. I believe it was their natural reaction to speak to me about what they were seeing, as they had done so many times before, and it took some time before they adjusted.

Daria's think-aloud sessions. As a participant, Daria often read the work of her students ahead of time and made notes. I imagine this was because the scheduling of time slots for think-aloud sessions were not lined up with when she needed to have read students' work for use in class. My guess is she needed to read the pieces before our sessions, so she took notes as she did. However, since she had these notes, and in the interest of moving more quickly—as think-aloud sessions can increase a teacher's reading time, a fact about which the consent form informed her—Daria had a habit of turning to her notes on the paper and reading from her notes rather than reading the paper itself and thinking-aloud directly in the session. This happened in both diagnostic and paper sessions, and on the surface it seemed to lessen when I reminded her on initial students' essays that she should read and think-aloud on the essays directly. However, the issue never really went away completely, as I could see her turn to her notes or review them before reading. Further, upon analysis, it was clear that even when appearing to

read and think directly, Daria was more frequently refraining from reading her notes until after she had read some of the student's essay directly, a characteristic of her think-aloud protocols that became evident when analysis revealed that she made observations about a student that did not connect to earlier statements made while reading, when there were hesitations between statements (signaling a change from reading to notes), and when she changed her mind about a student's fit to the course on either side of these hesitations (signaling that what was in the notes contradicted her reaction in that moment slightly).

Hannah and Terence's think-aloud sessions. Both Hannah and Terence were new to think-aloud protocols, so they experienced some difficulty verbalizing while reading at first. Hannah had the habit of reading, then reacting at the end, but we worked on thinking-aloud through training and early essays so she could realize that I wanted her to verbalize all thoughts, even those thoughts formed in the process of understanding a student's essay. She improved as we proceeded. The same was true of Terence. Terence started out saying almost nothing in training, other than her reported conclusion at the end, but as we completed the think-aloud training and she proceeded to read, she got better at thinking out loud.

Specifically, I realized two things were affecting Terence's ability to think-aloud, and these may well have been factors for Hannah as well. First, Terence had the mistaken belief that I wanted her put-together "thoughts" rather than all her thoughts in the process of thinking, a mistaken assumption I corrected by reviewing the directions with her when she said very little during training (Terence, Diagnostic Session). Second, she was fearful, particularly as we were training on a sample essay, that she was being tested to see if she got the "right" answer or she was being tricked by a purposefully difficult essay

to assess (Diagnostic Session). I assured her that the goal of the research was not to test her, but instead to learn from her expertise on the course what she focused on while reading Course 11 students' work, and that I had chosen the essay from previous years as only a sample for us to train on before teachers started with their own students' essays. Once these fears were quieted, she improved going forward. I do believe, in Terence's case though, that the hesitation may never have fully receded since when she was thinking aloud and placing students on the adequacy survey, she flagged far fewer students as atypical of the course than the average of other teachers. It is possible that she was nervous about the scrutiny, and that may have impacted her thinking-aloud.

Janet's think-aloud sessions. During the diagnostic think-aloud sessions, Janet was hesitant at first to verbalize, like many others, but through training, she became familiar with the process and she settled in. However, when we met later in the semester for the paper think-aloud sessions, while she didn't struggle to verbalize, she was far more distractible, her thinking was harder to follow as she would jump from thought to thought, and instances of off-topic chatter and interruptions increased. In a side comment, she revealed that her daughter was getting married in a week or so, and she was in the thick of wedding preparation, and I believe that fact accounts for the change in her think-aloud sessions.

Odysseus's think-aloud sessions. Odysseus took right to the process of thinking-aloud, and the early hesitation I felt with all the other participants, a natural response to feeling scrutinized, was interestingly not a problem at all for Odysseus. There were a few students on which Odysseus hesitated to reveal his thoughts immediately. Instead, he verbalized that he was checking the next paragraph, but often did not say what he was

checking for in the paragraph. However, in the few instances where this happened, Odysseus would begin thinking again before I could even remind him that he should think aloud about what he was checking for. One other minor issue was noted in the think-aloud sessions. As he was feeling under the weather at one point in the semester, having lost a little of his voice due to a cold and cheering at a baseball game, we stopped a few times for him to drink and soothe his throat. In these in-between moments, we would turn to some off-topic chatting, almost more so on my part, but when he resumed reading, he would get right back to work.

Data Analysis Limits

The first limit to data analysis is inherent in the amount of data collected. It is quite difficult to analyze in-depth when so much is collected. I tried to address that issue by choosing to forgo analysis of the paper think-alouds collected, as the research questions could be answered without them. While they will of course provide additional understandings, those are understandings that are not lost, as this data can be analyzed later. A second limit has to do with my history teaching Cardinal College's courses, reading placement, and working at the college as an adjunct instructor since 2004; I have my own understanding of the courses, my own thoughts on student work, and while I have worked hard to acknowledge and account for it, my own viewpoints have likely influenced how I understand the data.

Data Analysis Strengths

Since I couldn't help but have an opinion—on the courses, on the placement essays, on the prompts and assignments, and on students' work—I kept lists and notes on my opinions during data collection so I was consciously aware of my feelings in the

moment. That way, I could consciously track if my own feelings were effecting how I saw the data. Additionally, I was careful during data analysis to build codes directly from the words and ideas of the participants themselves as much as possible.

Finally, during the coding stage, I frequently had coded samples of transcripts rechecked by two colleagues who were not associated with the study, but who had background in teaching and tutoring writing. I asked these colleagues to see if my coding made sense and to problematize and raise questions about the coding. This check allowed me to rethink coding decisions in ways that mitigated the influence of my own interpretations. Let me offer one story on a code my colleagues problematized as an example. During an earlier stage of coding, I had coded places where placement readers noted that a student had copied from the text word for word both as a comment on how the student worked with text and a comment on the students' reading. One of my colleagues flagged this coding as potentially problematic. She asked what in the transcript I was basing this decision on. Through her questioning, I realized that I had noted an instance where a placement reader made this connection, and since I had a lot of experience with students who had trouble reading and would resort to recopying when they didn't understand, I made a generalization. Based on my colleague's challenge, I was able to recode the transcripts more in tune with the meaning and intention of the participants in particular moments.

Conclusion

By looking first at the moment of placement; then at the first year course (both the department and teachers' curriculum and course design, as well as the teacher's perception of the student's work, development, and performance), the data collected for

this study was gathered to validate a key assumption of the interpretive argument of Cardinal College's writing placement program—the use of its teachers' expertise to place in ways aligned with teaching and learning goals. This chapter has detailed how the study was designed and conducted to address this assumption. Chapter Four will go on to present the results and findings of the data analysis.

It is important to note again that conducting a validation argument for the program will take far more than this study and its research question, but this study offers a beginning to that work. The description and discussion of the context of Cardinal College, and the investigation of a key assumption of the interpretive argument of the program allows the program to study itself—to look closely at how the criteria under discussion by the placement raters and the teachers in the program relate and align—through which the program begins to check the effectiveness of its procedures and either revise or plan future studies to help make an argument for the validity of its decisions.

While this study is certainly helpful to Cardinal College, additionally, this case can be instructive for a larger audience. Those who share similar teaching and learning contexts will find the description and context of Cardinal College offered in this chapter helpful for determining the relevance of this study to their own situations. Regardless of a clear relationship among Cardinal College's and other college's contexts, the study of specific sites like this one, with published results that investigate the arguments for the validity of the decisions being made, can help in the creation and promotion of a culture in composition where writing assessment is done with a focus on teaching and learning.

CHAPTER FOUR

DATA ANALYSIS RESULTS AND FINDINGS

As outlined in previous chapters, the focus of this study overall was a beginning inquiry into the validity of the placement process at Cardinal College (a pseudonym)—a small, private college in the suburbs of a mid-sized, Mid-Atlantic city. Cardinal College's writing program uses Smith's (1992/2009, 1993) *expert reader model* to place students in one of two reading and writing integrated courses. To recap previous explanations, in Smith's model, an expert rater is one who has recently taught the course for which they are making placement decisions, and in his findings, these raters more reliably and adequately make placement decisions by employing their teaching and learning expertise to look not only at students' writing, but at the undertaking of teaching these students. Given this focus, this study considered the following research question:

- Do expert readers at Cardinal College use curriculum/course related and teaching specific criteria to offer a pedagogically useful and purposeful, view of students' placement essays? Are expert readers' criteria (and by extension their decisions) linked to course criteria?

Pursuing this validity-focused research question involved, as Chapter Three made clear, two different pieces of work: outlining the proposed interpretation and use of a test and evaluating the proposed interpretation (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Kane, 2006), or in Kane's (2006) terms specifically, the work involves sketching the *interpretive argument* and the *validity argument* (p. 22).

Chapter Three specified the writing program's interpretive argument, and through an examination of this interpretive argument, it became apparent that the key assumption on which the adequacy of the program's placement process rests—that expert placement readers chosen because of their recent course teaching experience use course-related and teaching specific criteria to make their placement decisions—had never been tested through formal study. Therefore, this study investigated expert readers' criteria and if and how these criteria were course related and teaching specific.

While the work of Chapter Three was primarily to sketch the interpretive argument and data collection rationale and to detail the steps of data collection and analysis, the work of this chapter focuses on the validity argument (Kane, 2006). More precisely, this chapter analyzes the collected data to tease out where the data support the interpretive argument and where the data reveal potential flaws, concerns, or questions that require rethinking the interpretive argument or conducting additional studies. In doing so, this chapter seeks to investigate the data toward some answer to the research question.

Chapter Four: Brief Overview

To determine if expert placement readers use course-related and teaching-specific criteria, data on readers' criteria were collected, namely think-aloud protocols taken as two placement readers, Adele and Theresa, read 26 placement essays and made decisions regarding two courses, Course 10 and Course 11, for the fall semester of 2013.

Additionally, the writing program records noting the decision counts for all original placements were collected. Due to the low number of student participants in the original placement sessions, additional protocols were collected as the readers reread 108

placement essays and remade placement decisions near the conclusion of the fall 2013 semester.

Data on the two courses and seven participating course teachers' criteria were collected in the form of course documents, interviews, a survey noting each students' fit-to-the-course, think-aloud protocols where course teachers read diagnostic essays rechecking the placement of students in the course, and lists of students' course outcomes and final grades. There were three Course 10 teachers (Daria, Adele, and Hannah), three Course 11 teachers (Terence, Odysseus, and Janet), one teacher who taught both courses (Theresa), and 118 students in total (59 from Course 10 and 59 from Course 11) consenting to the study.

The primary focus of analysis for this study was on coding and analyzing the placement and diagnostic think-aloud transcripts for specific criteria of reading, writing, and thinking under discussion for Course 10 and Course 11. Additional understandings regarding criteria, however, were gained by examining the surveys, course documents, and interviews with teachers, as well as the final course results. Finally, since Smith (1992/2009) specifically discussed that expert placement readers may also note behaviors of "the writer as a student" as opposed to just the characteristics of the essay—which Smith claimed is *privileged course knowledge* that helps expert readers make their placement decisions (p. 183)—the placement transcripts were also coded for these statements.

An examination of these data sets individually and a comparison or combination of some of these data sets allowed for four subquestions inherent in the research question to be answered:

1. What are expert-readers' placement criteria?
2. Are expert-readers' placement criteria teaching specific?
3. Are expert-readers' placement criteria linked to course criteria?
4. Are expert-readers' placement concerns and decisions linked to course concerns and decisions?

Going forward, I answer each subquestion in turn by considering and comparing relevant data and presenting the resulting conclusions along the way. In addition, where relevant, I situate the results with respect to relevant research discussed and detailed in the literature review of Chapter Two.

Specifically, I first detail expert placement readers' Course 10 criteria, Course 11 criteria, and what I call the edge criteria—that is the criteria that represent the border between these courses, gleaned from an analysis of students who presented placement difficulty. Included in this detailing is a discussion of students' reading and their work with text as key criteria that seem to help placement readers differentiate between students they believe are better served by one course over the other.

Next, I argue expert readers' placement criteria are teaching specific as readers note and use what I have determined are two different types of what Smith (1992/2009) terms privileged course knowledge in making their placement determinations. Then, I claim these placement criteria are course related as the criteria help teachers determine placement of difficult students. But I also note that two other types of reader comments—where students are noted as being on the boundary of the course and where the Course 11 reader makes a teacher recommendation—expose problems with Cardinal College's use

of the expert reader model as expertise on an old course that existed between the two existing courses still influences placement decisions.

Comparing placement criteria to course criteria, I show Course 11 teachers' criteria differ significantly from each other and from the placement criteria, and I discuss how these differences are likely caused by the population of students that individual teachers have experienced, both in the past and presently. With Course 10, I show the consistency in teacher's criteria, and I argue that this consistency can be potentially useful for helping the writing program understand the most concerning habits of reading, a necessary understanding if the program is to redistribute the old Course 11A student population into the two existing courses.

Finally, I use all data sets to track problematic and interesting sets of students through both placement and the course. Here I argue that the problematic students reveal a clear relationship between Course 10 placement and Course 10, but again expose a disconnect between Course 11 placement and Course 11. However, I assert that regardless of the Course 11 placement-to-course misalignment, the designations and decisions made at placement hold some relationship to student course grades and outcomes. This examination shows that the influence of the old course, while problematic, also contains important understandings that must be considered moving forward. Further, this examination exposes the difficulty of deciding problematic student placement as problematic students face possible demotivation if they are too far above the lower course and possible struggle if they are too far below the higher course. Overall, this chapter argues that expert readers do use course-related and teaching specific criteria to make placement decisions that are relevant to teacher's course work

and teaching decisions, but not fully. Changes in the placement reading process and the courses must be made and studied to ensure the relevancy of placement decisions to the courses, particularly for problematic students.

What Are Expert-Readers' Placement Criteria?

The first subquestion of the research question involves determining what criteria expert placement readers use for each course since knowing these criteria is necessary for consideration of all the other parts of this research question. To uncover the readers' placement criteria, think-aloud protocols were collected at two different times: first, as readers read placement essays over the summer during three separate placement sessions in August 2013, and again when readers reread placement essays in three redo sessions scheduled in November and December of 2013.

This section discusses data from the think aloud transcripts of these two sets of placement reading sessions. It is necessary then to begin with a review of some information from earlier chapters that will contextualize this data and its analysis. I start by reviewing the placement-reading process at Cardinal College. Then, I offer an overview of the placement sessions, more specifically a summary of the numbers of participants and decisions made. Next, I give a description of the reading and writing criteria that emerged from coding and analysis of the transcripts overall, before providing important notes on how I coded and grouped the data for analysis that will be relevant to how I've organized my findings in this section. Once this context is established, I proceed to present the discussion of the data and results of my analysis.

Review of the Placement Reading Process at Cardinal College

It will help to review what the placement-reading process is like at Cardinal College to establish a context for the comments made subsequently as the data is discussed. Both think-aloud recording sessions for original placement and the redo sessions followed the established process of placement reading as I described it in Chapter Three. To recap, at Cardinal College, there are two placement readers, one for each of two courses, Course 10 and Course 11, and these readers must have recent course teaching experience in the course for which they are reading (Smith 1992/2009, 1993). All placement essays are read by the Course 11 reader first, as that is the most populated and general course (Haswell & Wyche-Smith, 1994), and this reader decides if students are in her course or not in. Those essays that the Course 11 reader pushes down are then read by the Course 10 reader, who decides if the student is in or not in her course. Any essays that are *tweened* (Smith 1992/2009, 1993)—pushed down by the Course 11 reader and up by the Course 10 reader—are then discussed, and a decision on the best placement is made by the readers in conversation with one another.

Overview of Placement Reading Sessions

Twenty-six participating students' essays were analyzed from think-aloud protocols collected during the three original placement sessions held over the summer. Twelve of these students were placed into Course 10, and 14 were placed into Course 11. Eight of the 26 students were described by the readers as Course 11A students (2 who were placed in Course 10 and 6 who were placed in Course 11). Course 11A, as I described in Chapter Three, was an old four-credit course that used to exist between the six-credit Course 10 and three-credit Course 11. Course 11A was discontinued by the

administration due to space constraints, which made scheduling the four-credit class difficult. Also, one of the participating students was a tweener in an original session—a student who was pushed down by Theresa, the Course 11 reader, and up by Adele, the Course 10 reader. The ultimate decision was for the student to be placed in Course 11 (see Appendix O for a table summarizing all 26 original placement participants and their placement decisions).

One-hundred-and-eight participating students' essays were analyzed from think-aloud protocols collected during three redo placement sessions. Originally, 56 of these students were placed into Course 10 and 52 were placed into Course 11. In the redo sessions, however, 44 students were placed into Course 10 and 64 were placed into Course 11, with different decisions being made on 18 students' essays (15 who were switched from Course 10 to Course 11 and three who were switched from Course 11 to Course 10). Thirty-four of the 108 students were now described by the readers as Course 11A students (15 who were placed in Course 10 and 19 who were placed in Course 11). Also, eight of the participating students were now tweened in these sessions with six of these students put into Course 10 and two of them put into Course 11 (see Appendix P for a table summarizing all 108 participants and their placement decisions in these redo sessions).

Description of All Reading and Writing Criteria

As all think-aloud transcripts were coded, 14 criteria of reading or writing that readers discussed on students' essays emerged in total. Table 2 lists and describes these criteria. Of the 14 listed criteria, all of the criteria except criterion number 13, writing process, were discussed in relation to placement essays. These criteria were developed

Table 2

Reading and Writing Criteria

Number	Name of Theme	Description of Theme
1	Language	Writers' word choice, vocabulary, & syntax
2	Grammar	Writers' grammar, punctuation, & spelling
3	Academic rhetorical conventions	Writers' naming the author, naming the texts, & audience awareness
4	Thesis/project	Writers' viewpoint, position or argument overall
5	Framing/moves	Writers' moves, framing, or paragraph/section purpose
6	Essay/ paragraphing knowledge	Writers' understanding of essay structure, paragraph structure, & paragraph breaks
7	Length/amount of writing	Overall length of the response as well as amount of quote versus writers' content/words
8	Working with text	How the writer works with quotes—including using, choosing, marking, setting up, & explaining quotes
9	Reading issues/quality	Writers' understanding of the texts
10	Assignment task	Writers' understanding or performance of the task called for by the assignment
11	Use of required texts	Writers' use of the required materials called for by the assignment
12	Print conventions	Writers' handwriting, indentation, & margins
13	Writing process	Writers' knowledge of the writing process
14	Thinking	Writers' ability to think about, extend, or connect with ideas from the reading or on the topic

directly from the language of the placement and course readers, so a little explanation may be necessary to detail and distinguish the criteria. When readers discussed students' essays, they would occasionally comment on a students' language. Here they were not interested in discussing the rules of language—grammar, mechanics, and punctuation. Instead, they were focused on choices related to wording, style, and vocabulary, including use of vague pronouns like “this” or “it.” This criterion is not to be confused with comments on grammar—most notably punctuation and spelling, but also fragments and run-ons. Academic rhetorical conventions was the term used by both placement

readers regarding conventions related to introducing authors and texts, referring to an author by his/her last name in subsequent discussion, and even audience awareness in terms of providing necessary details and explanations.

Many of the teachers at Cardinal College use the term *project* to refer to the purpose, point, and work of an essay or paper. The choice of this term is meant to distinguish and differentiate a project from the term students are more familiar with, a thesis. To many teachers, the term project helps students understand that their papers must *do* something with the materials they have chosen to use, to arrive at new understandings, arguments, or knowledge, rather than passively discuss topics, ideas, or content. However, teachers are also aware that students more frequently label the point of their paper their thesis. As such, the teacher-readers used either term as they discuss students' essays.

If the argument overall was labeled the thesis or project, the readers used the terms *framing* or *moves* to note a writer's purpose in a moment or section of their text, and they frequently commented on their ability or inability to determine a student's purpose in a moment, as well as the sequence and planning of moves. The placement readers seemed to see moves and framing as distinct from essay knowledge or paragraphing knowledge more generally. Regarding this criteria, readers commented on the sense they had about a students' familiarity with and knowledge of classic essay and paragraph structure—introductions, titles, conclusions—as well as their understanding of when paragraph breaks are called for.

The criterion of length is fairly self-explanatory. Readers often noted if a response was short or lengthy, but they also noted, as part of length, when they felt that the student

had written very little themselves in comparison to how much they had quoted. On the subject of quotations, the majority of the comments on the criterion of working with the text included comments on quotations. Frequently, readers praised the use of quotations or noted problems with marking quotations (text boundary issues), but they also commented on the choice of useful or unhelpful quotations and the introduction and explanation of quotations. Next, readers commented on how much of an assigned text students understood, including comments on localized misreading versus overall misunderstanding. Understanding or misunderstanding was also a theme of readers' assignment task comments. Here readers noted when students seemed to misunderstand the assignment or substitute a different task in its place.

While related, I separated assignment comments from comments on the students' direct use of both of the required texts provided with the prompt. Using the required texts was part of the assignment, but placement readers distinguished between students who understood they should engage directly with these texts, acknowledging their voices, and students who interpreted one or both of the provided texts as information or topics. Thus, it was possible for readers to feel a student was attempting the assignment, but not directly using both texts, and it was possible, and much more frequent, for readers to feel a student was not doing the assigned task even though he or she had writing directly related to both texts.

Occasionally, several conventions associated with print were discussed by readers such as maintaining straight margins, the indentation of paragraphs, and neat or messy handwriting. Also, a few times students' knowledge of the writing process was discussed as evidence of planning or editing was noted. Finally, readers commented on students'

thinking, specifically the students' attempts to relate to, extend, or connect ideas from the reading or other ways of thinking on the topic.

Some Notes on Coding and Analysis

It is important as well to note some other coding elements that will be relevant in the discussion on this subquestion going forward. First, as placement readers discussed these 14 criteria, their discussion would frequently fall on a continuum of ability with the comment often having a negative or positive connotation, not unlike Huot (1988) found with his raters' comments. For example, a reader could discuss reading quality negatively, noting misreading, as Adele did on student 1010-01-07 in Redo Session 1, saying "Okay, this is a misreading of the text here." Or, a reader could discuss reading quality in positive terms, noting an important idea understood, as Theresa did on student 1011-03-09 in Redo Session 1, saying the student "understands that Graff is using an analogy" and that the student "gets what Graff is saying about...things being hidden." The positive, negative, or mixed nature of comments made on each student's placement was noted.

Second readers would often comment on aspects of these criteria that could be considered subcategories of the criteria (most notably with codes 9 and 10 in Table 2, reading issues/quality and assignment task). For example, while the majority of discussions on reading quality named the issue broadly as *misreading*, on occasion the particular way of misreading was explained further. For example, in Redo Session 1, Theresa described student 1010-05-04's misreading as "making things up" to emphasize the way the student seemed to follow a line of thought not in the text, although seemingly related, but attributed to the text. Subcategories of comments were noted as I coded.

Third, once I finished coding, I counted the number of students in original sessions and the number of students in redo sessions on which each criterion was discussed by each reader, as well as if the discussion of a criterion was negative, positive, or mixed overall. Additionally, I counted the total number of students on which each criterion was discussed overall with both sessions and both readers factored into one count. Since many students were discussed twice (once in the original session and once in the redo session), and some were discussed as many as four times (by both readers in both original and redo sessions), the total count was calculated by removing any instances where the same student was discussed on a particular criterion by both readers or in both sessions. These counts allowed me to see which criteria were more or less important to making course placement decisions and to characterize the discussion of those criteria. More will be said on these issues as I discuss and quote specific comments from the data as they are relevant to my findings.

Finally, after finishing coding and moving to analysis, I made the decision to separate the participating placement students into two groups based on noted placement designations for analysis. First, I grouped together what I called the unproblematic students. In two separate reading sessions, some students were cleanly and consistently placed. Given Smith's (1992/2009, 1993) findings that disagreement about placement decisions arises because some students do not fit neatly into course options, I believed these unproblematically placed students would make good representatives of the criteria for their corresponding courses (see Appendix Q for a table listing all 52 unproblematic placement students, including if they participated in only the redo session or both sessions). Second, I grouped together what I called problematic students—students who

were noted as falling between the courses (called Course 11A students), students who were tweeners in a session (called tweeners), and students who received a different placement decision in the redo session than their actual course placement (called different decisions). Again, drawing from Smith's findings, I believed these students could be examined to better understand the border or space between the two courses (see Appendix R for a table listing all 56 problematic placement students, including if they participated in only the redo session or both sessions). In what follows, I discuss the results of my analysis of the placement criteria. Using the placement think-aloud transcripts—from both the original and the redo sessions—I detail expert-readers' placement criteria, noting Course 10 and Course 11 criteria, as well as what I am calling the edge criteria. The unproblematic students were used to determine the criteria of the courses while the problematic students were used to discuss the edge or space between these courses.

Overview of Placement Think-Aloud Analysis Results

Going forward, I present the results of my analysis of the placement transcripts from both the original and redo placement sessions. First, I look at the placement criteria overall, including the combined total number of students for which each criterion was discussed by at least one reader, and a brief explanation of the major or key criteria with examples. Second, I detail and discuss the Course 11 criteria followed by the Course 10 criteria. This presentation reveals how overall many of the same criteria are key for Course 11 as are for Course 10, with differences in comments related to a continuum of competency on specific reading and writing issues. Course 10 students typically receive negative comments on the criteria while Course 11 students typically receive positive

comments or wording that suggests a less severe position on the developmental continuum. I also discuss some interesting differences between Course 10 and Course 11, showing that the frequency of comments on certain criteria, grammar and framing/moves, varies greatly between Course 10 and Course 11. Third, I discuss the edge students. Specifically, I argue that these students show more of a mix of positive and negative comments on each criterion, which indicates that they have competency on some aspects of a criterion and less on others. And I show that in some ways, these students do not show the same competencies as their unproblematic course counterparts, but where they appear to, an analysis of what readers say indicates less certainty on their assessment and more weighing of conflicting combinations of criteria than with unproblematic students.

Placement Criteria Overall

To get a sense of which criteria mattered most to placement, I counted how many of the total 108 placement students had a particular criterion mentioned in the discussion of their essay. As I noted earlier, many students were discussed twice (once in the original session and once in the redo session), and some were discussed as many as four times (by both readers in both original and redo sessions). Therefore, the total count was calculated by ignoring instances where the same student had the same criterion mentioned by each reader or in multiple sessions by different readers. Table 3 presents the total number of students discussed for each of the 14 reading and writing criteria with repetitions removed.

When the discussion by both readers on all 108 students in the original and redo sessions is accounted for, the most important criteria for placement overall—regardless of session, course, or problematic/unproblematic designation—are the students' quality of

Table 3

Total Number of Students Discussed for Each Placement Criterion

# of Students	% of Students	Name of Criterion
105/108	97.22%	Reading issues/quality
96/108	88.89%	Assignment task
82/108	75.93%	Essay/ paragraphing knowledge
76/108	70.37%	Working with text
43/108	39.81%	Framing/moves
32/108	29.63%	Length/amount of writing
29/108	26.85%	Language
29/108	26.85%	Grammar
22/108	20.37%	Thinking
21/108	19.44%	Use of required texts
17/108	15.74%	Academic rhetorical conventions
13/108	12.04%	Print conventions
5/108	4.63%	Thesis/project
0/108	0%	Writing process

reading, their understanding of the assigned task, their knowledge of paragraphing or essay structure, and how they work with the text (specifically text boundaries and the explanation of quotations). Several other criteria are also factors in placement, but to a lesser degree. Before moving on to discuss the criteria of the two courses separately, let me first offer some examples to detail the four major criteria overall since, as I show going forward, these four major criteria are essential to all courses and student groups.

Reading issues and quality. The most frequently discussed criterion looked at students' understanding of the required readings associated with the placement prompt. The placement prompt asks students to read a short excerpt from Gerald Graff's "Hidden Intellectualism" and a newspaper article by Pete Zapadka on the demotion of Pluto from a planet to a dwarf planet (see Appendix A for the full placement prompt and readings). As was discussed extensively in Chapter Three, Cardinal College's first-year writing courses integrate reading and writing, so it is not surprising that reading quality would

come up as a chief concern of placement readers. Regardless of placement session or reader, reading quality was discussed again and again. What follows are a few common occurrences in the data in regards to reading issues and quality.

Most frequently, readers would note three aspects of reading quality: specific moments of misreading, overall misreading of a text or texts, or important ideas from a text understood by the student. For example, as Theresa discussed student 1011-00-01 in Original Placement Session 6, she noted a key idea from the text that the student understood, saying the student “understands that Graff is making a comparison between the real intellectual world and the world of sports” which was “a pretty good indicator that this person is a [Course 11].” Yet, as Theresa discussed student 1010-05-03 in Redo Session 3, she noted a specific moment of misreading in regards to Graff’s text where the student “thinks what’s kept hidden is...current events.” While many students had specific moments of misreading or moments of strong reading discussed about their essays, it was also common for readers to make an overall statement on reading quality, typically negative, such as with Adele’s discussion of student 1010-04-06 in Redo Session 3, which characterized the student’s piece as a “total misread” of Graff, focusing on sports, and as having “nothing to do with either reading.”

In addition to statements on misreading and strong reading, whether generally or overall, readers also categorized certain reading issues into several subcategories, most notably “fake reading” (Adele, Original Session 7, 1010-02-01), “making things up” (Adele, Redo Session 2, 1010-01-05), “approximation” (Theresa, Redo Session 2, 1010-04-09), and “substituting... clichés” (Theresa, Redo Session 1, 1010-01-04). Each of these categories suggested misreadings that resulted from not closely paying attention to and

using the text. With *fake reading* and *making things up*, the suggestion was the student's reading was a fraud. Either the student attempted to use words or phrases from the assignment or texts themselves to give the appearance of understanding, when the student hadn't really understood, or they have invented ideas to suggest they read the text fully, when they hadn't. *Approximation* suggested that the student was writing about the reading while relying on their memory of what they'd read rather than the text directly, and the label of *substituting clichés* indicated the student didn't pay attention to the actual words of the text and rather let their assumptions about the subject or topic mislead them.

Assignment task. The next most frequently commented on criterion for placement readers was how well a student understood and addressed the assigned task of the placement prompt. Thus, a brief description of what the placement prompt asks for is necessary here. Cardinal College's placement prompt asks students to write a single essay in which they "explain [their] understanding of what Graff calls the 'real intellectual world'" and "go on to apply it to...the demotion of Pluto." The prompt asks them, as a concluding question, "How might Graff's ideas help you understand what the big deal is about 'demoting' Pluto?" In short, the prompt asks students to use Graff's text as a lens to look at the demotion of Pluto.

As readers thought-aloud on students' responses to this prompt, their comments could be categorized into one of three main types in regards to the assignment. Frequently, they discussed students who were not addressing the assignment at all, at times suggesting that students misunderstood what the prompt was asking them to do (Theresa, Original Session 5, 1010-07-06) or didn't know how to do what the prompt asked so they substituted another task (Theresa, Redo Session 1, 1010-04-08). In

addition, placement readers also noted students who successfully addressed the assigned task (Theresa, Original Session 5, 1011-03-13) and students who at least attempted the task, albeit with the implication that their attempt was less than successful (Adele, Redo Session 2, 1010-05-05).

Students who misunderstood or changed the task would often write one of a few kinds of responses instead of the one asked of them. First, many would write a summary of one of the texts (Adele, Redo Session 3, 1010-07-04) or a summary of both back to back with little or no connection between them (Adele, Redo Session 1, 1010-06-01). Second, others would write an essay that seemed to start with the task but would quickly spring off into a personal experience essay (Theresa, Redo Session 1, 1010-01-01) or would express the student's opinion or reaction to the argument or details of one of the texts (Theresa, Redo Session 3, 1010-04-06). There was one other occurrence that placement readers noted in regards to students' ability to write in response to the prompt. While the prompt calls for one essay to be written, at times students would produce two separate essays in response to the prompt. These two essays could be completely separate from each other, each part dealing with one text in isolation from the other, or the second essay could make some connections to ideas in the first, despite being on a separate sheet of paper or under a new essay title or number (Adele, Redo Session 1, 1011-08-01).

Essay and paragraphing knowledge. On the issue of essay and paragraphing knowledge, the third most frequently made comment by placement readers, readers were concerned with how the students' responses showed their understanding of classic essay structure (an introduction that set up a point, body paragraphs explaining and supporting the point, and a conclusion) and, failing that, how the students' responses demonstrated

paragraphing knowledge like breaking to new paragraphs for new ideas or points. Since essay knowledge and paragraphing knowledge were often discussed together, I did not see these as separate criteria, but rather aspects of the continuum of knowledge readers were looking for in students' essays. On one end of the spectrum, readers commented on students who were demonstrating a full understanding of classic essay structure, and by extension paragraphing structure (Theresa, Original Session 6, 1011-09-08), and at the other, they noted students who wrote responses that were one long paragraph with no paragraph breaks or the separate elements of an essay (Adele, Redo Session 3, 1010-06-06). It was also possible to fall somewhere in the middle with a kind of two or three paragraph response that lacked the formal aspects of the classic essay, but would make separate points in several paragraphs (Theresa, Redo Session 3, 1010-04-03).

Working with text. The fourth most frequently commented on criteria by placement readers had to do with how a student worked with the text or texts. While readers would comment on the quality of moments of paraphrasing or summary on occasion (Theresa, Redo Session 1, 1011-08-07), their concern was more frequently with how students used and treated quotations. Sometimes placement readers commented on the choice of quote—typically they praised the student as choosing a good quote or moment (Theresa, Redo Session 2, 1010-01-08). Much more frequently though, readers were concerned with how students marked the boundaries of quotes and how they explained and integrated quotes into their points.

When commenting on quotation boundaries, readers' comments would often reveal their judgment whether the student's copying could be easily fixed by teaching them conventions for marking quotes or if there were larger issues with low language

proficiency or reading issues. The judgment here usually had to do with how much unmarked quotation the reader noticed. For example, as Adele discussed student 1011-03-01 in Redo Session 2, she noted the student copied a quote without quotation marks, saying “You can actually see they are like lifting word for word,” but very quickly she moved on to discuss the student’s understanding of the text and never mentioned copying as an issue again. However, with her discussion of student 1010-06-13 in Redo Session 1, Adele mentioned copying several times, and she thought the copying signaled the student didn’t really understand at all as she said the student “[pulls] little words [from the text] to make it sound good, but it just doesn't make sense.”

On the explanation and integration of quotations, readers frequently discussed the need for students to *unpack* their ideas. Comments on unpacking signaled the reader felt the student didn’t work enough with the text directly to fully come to a clear sense of what they were saying themselves or to show the reader what they were saying. For example, as Adele discussed student 1011-08-01 in Redo Session 1, what she said revealed she thought the student had not done enough to show their interpretation of the text. She said, after reading the student’s explanation of a quotation, “But they don’t say what those intellectual skills are. So it’s like a lot of quotation here, but they’re not doing enough work with the quotation.”

While placement readers could often determine if they felt the student needed to unpack for their own understanding, for their reader’s, or for some mix, readers also commented on unpacking when they couldn’t decide if the student truly understood or if they just didn’t know to show their interpretation in writing. For example, as Theresa discussed student 1010-01-11 in Original Session 6, she said the student had written a

response that was like “a fill-in-the-blank paper” or “end of the chapter” answer without explanation or unpacking. As she continued, explaining why she will push it down for a second reading, she said:

Now my thinking is that with the right teacher, that person would be told you can’t do that. You have to spell this out. Um, my issue is whether this person needs so much practice in that and has so little understanding of sort of the larger intellectual moves that maybe [Course 10] would be better. (Theresa, Original Session 6)

In short, she couldn’t decide if this student would easily add the explanation and work with text, or if being asked to do so would cause struggle.

Course 11 Placement Criteria

Now that I have described the placement criteria overall, in particular the four main reading and writing criteria discussed at placement, I want to move to detail the criteria of the two courses separately so I can go on to compare the criteria of the two courses to each other. Starting with Course 11, I discuss the criteria determined from an analysis of Theresa’s original and redo session think-aloud protocols on all unproblematic Course 11 student essays, showing her criteria overall and trends in the frequency of positive versus negative comments. Then, I discuss the most important criteria for Course 11 placement, making sure to contrast this discussion with the earlier comments on placement criteria overall before finishing up with a discussion of those students on whom I have both original and redo placement data.

Criteria overall. As I analyzed Theresa’s original and redo placement session think-aloud protocols, I focused on unproblematic Course 11 students—those 26

participating students who were placed into Course 11 in both sessions and who were easily placed without being designated as tweeners or flagged as old Course 11A students. As I noted earlier, I counted how many students Theresa discussed for each criterion and if that discussion was positive, negative, or mixed. Table 4 details the unproblematic Course 11 students' totals as discussed by Theresa for each of the reading and writing criteria.

Table 4

Total Number of Unproblematic Course 11 Students Discussed by Theresa for Each Placement Criterion

# of Students	% of Students	Name of Criterion
26/26	100%	Reading issues/quality
25/26	96.15%	Assignment task
18/26	69.23%	Essay/ paragraphing knowledge
15/26	57.69%	Framing/moves
14/26	53.85%	Working with text
5/26	19.23%	Language
3/26	11.54%	Length/amount of writing
3/26	11.54%	Thinking
3/26	11.54%	Print conventions
3/26	11.54%	Thesis/project
2/26	7.69%	Use of required texts
2/26	7.69%	Grammar
1/26	3.85%	Academic rhetorical conventions
0/26	0%	Writing process

Positive rather than negative comments. As Theresa discussed the unproblematic Course 11 students, the overwhelming majority of her comments were positive or mixed overall. Of the 26 unproblematic students, only seven students had a fully negative assessment on some criteria noted. The other 19 students had all positive, all mixed, or both positive and mixed comments. In addition, only two of the seven students had a fully negative assessment on more than one criterion. The other five

students had only one criterion with a fully negative assessment coupled with mixed or positive assessments on other criteria.

Important criteria for Course 11. Looking at the total number of students commented on for each criterion is a fairly easy way to determine the criteria most important for Course 11. The criteria with the highest total number of students are the following: reading issues and quality, assignment task, essay and paragraphing knowledge, framing and moves, and working with text. While this is one important way to determine those criteria that matter for Course 11 placement, four of these five criteria are already important for placement overall, and among the top criteria, only framing and moves changed position in the list for Course 11 placement. Therefore, we have to look a little deeper into the qualitative end of the data to see (a) what aspects of these particular criteria matter for Course 11 placement; (b) how the breakdown of positive, mixed, and negative comments plays out with specific criteria; and (c) what this all reveals about the relationships among Course 11 criteria involved in determining placement.

Reading issues and quality. For Course 11 students, Theresa's comments on reading typically were observations about positive or strong moments of reading or observations about both strong understandings mixed with moments of misreading. In fact, there was only one student who was discussed in this group with only negative comments on reading, student 1011-09-05, and even that student was not characterized as having misread overall. Instead, Theresa discussed how she understood what the student was "trying to say" about Graff at times, but the student also reduced Graff to "something to do with rules" (Redo Session 3). Therefore, it seems unproblematic Course 11

students, as a group, hardly ever exhibit more serious issues with misreading, at least as perceived by Theresa.

Reviewing the 26 unproblematic Course 11 students further reveals a few instances that demonstrate that Theresa expects that Course 11 will handle issues of misreading of a particular degree. For instance, on student 1011-08-04, Theresa said that the student “[didn’t] recognize that this was a debate” but she also said the student was “very close to having a pretty good reading (Redo Session 1). And about student 1011-03-05, she said, “they miss the whole point about the convening of the [demotion meeting], and the debate, and the astronomers” but “he could certainly do [Course 11]” (Redo Session 3). For Theresa, it seems, some misreading is expected from Course 11 students, and her assumption is that the course will help students work on these issues of misreading. Yet, her comments suggest that if students exhibited more severe misreading or exhibited misreading coupled with problems with some of the other main criteria, she would consider making a different decision. In fact, there is a hint that Theresa often weighs the criterion of reading against the criterion of essay structure as she makes decisions on placement, as I will show with student 1011-03-09 as I discuss comments on the student’s meeting the assignment task.

Assignment task. As with reading, the criterion of assignment task was discussed almost completely in positive terms. Eighteen of the 25 students commented on for this issue received positive comments as they attempted the assigned task and successfully addressed the assignment. Six of the 25 students received mixed comments, most commonly of the variety that indicated, as with student 1011-03-04, that the student “tries” to make a connection between the texts (Redo Session 2) or, as with student 1011-

09-09, that the student “doesn’t make strong connections” overall but at least “puts the major points...in connection to one another” (Redo Session 2). In short, as with reading, unproblematic Course 11 students typically have less serious assignment issues in that they complete or at least attempt the task with some success.

Only one of the 25 students received a fully negative assessment, and what was discussed about this student, 1011-03-09, suggests that the issue was the student “chickened out” and “didn’t make the connection at the end” (Redo Session 1). But additionally, what else Theresa noted with this student shows she weighed a negative view of the student’s attempt of the assignment against other criteria, like essay and paragraphing knowledge and reading, to decide a Course 11 placement. In addition to noting the lack of connection, Theresa said the student had set up a “fairly competent essay” structure and she said he “just [needed] to be pushed a little harder in his reading” to get the connection he lacked (Redo Session 1).

Essay and paragraphing knowledge. On the issue of essay and paragraphing knowledge, unproblematic Course 11 students were never commented on with fully negative assessments. Sixteen of the 18 students receiving comments on this issue had positive assessments, and looking at the transcripts of these readings shows that all 16 students had at least some paragraphing knowledge with most also exhibiting either a move to introduce and set-up their point or a full essay structure. The remaining two of 18 students received mixed comments, but looking more closely at these two students is revealing for determining at what point Theresa becomes concerned with students paragraphing and essay knowledge. Student 1011-03-05 was described as having “one full paragraph” and “a couple of little paragraphs” after that (Redo Session 3). Student

1011-05-02 only wrote one large paragraph, but Theresa said, “it could be divided into paragraphs quite easily” since separate moves seemed apparent inside the large paragraph (Redo Session 3)

Framing and moves. In regards to framing and moves, it is interesting that no unproblematic Course 11 student was negatively commented on for framing and moves. Instead 13 of the 15 students commented on for framing and moves had positive comments and two had mixed comments. As I already discussed above, when Theresa read student 1011-05-02, she observed that the student had moves, even though the paragraph breaks that usually go with them were not present (Redo Session 3). Pointing out the presence of moves was something frequently noted with Course 11 students, and even on occasion Theresa would list out verbally a student’s sequence of moves. Additionally though, with mixed comments, Theresa would signal some issues with moves and framing that were characteristic of Course 11 students. Student 1011-08-04 is representative as Theresa noted that the student “spends a little bit too much time summarizing” (Redo Session 1). Frequently, Theresa would note, as she did here, that a student had moves or framing in his or her essay, basically a positive, but she would also point out where those moves were ineffective as a negative.

Relationship of essay and paragraphing knowledge to framing and moves.

Returning to my earlier discussion of Theresa’s comment on essay and paragraphing knowledge, these comments suggest two important aspects of unproblematic Course 11 placement. First, her comments on student 1011-05-02 show a relationship of essay and paragraphing knowledge to moves and framing, as does the fact that 13 of the 15 students commented on for framing and moves were also commented on for essay and

paragraphing knowledge. For many writing teachers, the relationship of framing and moves to essay and paragraphing may be obvious. Since academic writers typically break paragraphs related to the point and purpose the writer wishes the paragraph to achieve, it is logical that there would be a relationship here.

But Theresa's comments on student 1011-05-02 not only suggest a relationship, they suggest that a student who has a sense of the moves required or a plan of moves for their paper, even without paragraph breaks or a full essay structure, is a student who is prepared for college level work. In essence, if Theresa can track and understand the reasoning for the student's points and planned discussion in the essay, then she believes that student understands the fundamental purpose behind paragraphing and essay structure, even if they do not yet have this structure. Furthermore, a student who understands framing and moves is more prepared than one who might have paragraphing and the classic essay structure alone, as this first student is focused on their point and purpose and lets that drive their decision-making more than a formula in their head that they are not quite sure how to make their own.

Working with the text. On working with text, unproblematic Course 11 students most frequently received comments on issues of unpacking, and they never received comments on issues of copying. This suggests that quotation boundaries are not an issue for Course 11 students, but their abilities to work closely with a text and its words are. In fact, of the 14 students commented on for working with text, 11 received a comment that related to unpacking. And, when the 11 students who received a comment related to unpacking or close reading were looked at more directly, 3 had positive comments praising some strong moments of close reading of a quotation, and 3 had negative

comments that suggested direct work with the text was necessary to explain a point from the text as the student lacked this direct textual work. The five students who had mixed comments either attempted to closely read a quote or quotes with some degree of success, or they relied on summary more than quotation.

For those students who summarized more than they quoted, the point seems to be that the student needed to learn to choose direct analysis and use of textual moments over a retelling of the text. There is even a suggestion that summary can take over a student's piece, as we can see with 1011-03-10 who had focused on strong ideas from the text but since "there's a lot of summarizing," Theresa waited to see that the student "[brought] Graff back into Pluto" (Redo Session 2). The same taking over of summary was true of 1011-08-04 who "quotes the right piece, a useful piece about Graff" but who "spends a little bit too much time summarizing the Pluto article" such that nothing gets done with the ideas from the earlier quote (Redo Session 1). In some ways here again, we see Theresa weighing a student's work with text in relation to other criteria. She seems to be considering, with both of the above students, as well as others, if the student's work with the text helps him or her develop a reading or response to the assignment and if it helps him or her use the text in purposeful ways, as part of a sequence of his or her own moves.

Students with both original and redo session data. There were five students in this group of unproblematic Course 11 students that were participating in both phases of study, and therefore have think-aloud data on their essays from both the original and redo placement sessions. Looking more closely at these five students, I investigate the ways Theresa's comments on these students were consistent with each other and how they changed across readings. Three of the five unproblematic Course 11 students with data

from both placement sessions had identical assessments made on three or more of the major placement criteria, with the only difference being one or two additional criteria included in one of the readings not noted in the other. These students were 1011-00-01, 1011-03-13, and 1011-09-03. One of the five students, 1011-09-02, had an identical assessment on three criteria, two of which were major Course 11 criteria, with a few additional criteria noted in both reading session that were not noted in the other session. The final student, 1011-09-11, only shared one criterion, the key criteria of framing and moves, between the two readings, with two additional criteria noted in the original session and four other criteria noted in the redo session.

As the first three students had all or a majority of positive comments, it seems that the differences between Theresa's separate readings on these students can be accounted for in part by the significant number of positive points for key Course 11 placement criteria that could be noted on these students. A closer look at the first four students, even the one with a few negative comments, reveals additional explanations: the differences between readings are part of the nature of differences in what criteria and in what order our minds attend to criteria on any given day, and the differences are connected to various ways course and teaching knowledge can be weighed and combined to help make a placement decision.

For example, with student 1011-09-03 in the original session, Theresa noted significant length and strong moves early on, but she also noticed a mix of positive aspects of reading with a moment of "a kind of misreading" where the student thinks Graff's "idea of hiding is literal" (Original Session 7). She claimed the reading issue is easily fixed as the course will teach the student to be a more careful reader, and thus she

made her placement decision. In the redo session, she again noted the strong moves but this time she noticed some close reading work with the text before coming to see potential problems with reading. Additionally, this time, she made the decision to place the student in the course, not by thinking about what the course will teach the student in regards to this one criterion, but by weighing the possible issue with reading against several other positive criteria she's noticed and by considering the nature of the on-demand task. She said, "I mean it's a misreading. It's interesting. It's not a deep and complex reading, but let's say given an hour and a half and the materials he had, it's an essay. He uses both texts" (Redo Session 2). The resulting judgment seems to be that, having only the minor reading issue, the student will be able to manage Course 11 without problems.

This example again shows that Theresa weighs key criteria against each other, particularly when she sees a mix of aspects on a key criterion. Thus, when she noted a reading issue, she knew she could keep looking elsewhere to weigh other key criteria against what she's already noted and see if any additional concerns emerge. Given that few other concerning issues emerged, the decision was relatively easy. The analysis of Theresa's comments on this student also suggests that in addition to weighing key criteria, Theresa also uses an understanding of the on-demand nature of the task and Course 11 knowledge to make the decision. I will consider the use of course knowledge more fully in later sections of this Chapter, but it is important to point out here.

Overall, the similarities in assessment between the multiple readings on these students, primarily on those criteria the quantitative count has already suggested are key Course 11 placement criteria, confirm the importance of these criteria for Course 11

placement. Further, the differences in assessment between the two readings of these students' essays do not seem to signal a problem. Instead, it seems plausible that there are different paths to a decision about placement generally, particularly for students who fit neatly in a course, and, as we know, thinking-aloud is not going to be a full reporting of what is happening in the mind (Ericcson & Simon, 1980, 1984). Sometimes Theresa mentally and verbally weighed various observed criteria against each other to make a placement decision, consciously acknowledging that she was thinking about course teaching expertise to help her make a decision; sometimes she weighed various observed criteria against each other to make a placement decision without being aware, or at least without verbalizing, which criteria weighed more heavily in her mind. It remains to be seen if differences in one reader's assessment on different occasions cause problems on other student essays.

Course 10 Placement Criteria

Turning to Course 10, I discuss the criteria determined from an analysis of all original and redo session think-aloud protocols on unproblematic Course 10 student essays. Since unproblematic Course 10 essays were read by both Theresa, as the Course 11 reader who pushed the essays down, and Adele, the Course 10 reader who decided their placement, it is important for me to eventually look at both readers' criteria. However, I begin by examining Adele's think-aloud protocols, as she was the Course 10 expert reader, showing her criteria overall and trends in the frequency of positive versus negative comments. Then, I use her transcripts to discuss the most important criteria for Course 10 placement, making sure to contrast these criteria with the earlier comments on placement criteria overall and for Course 11 before finishing up with a discussion of

those students on whom I have both original and redo placement data. As a final discussion on Course 10, I go on to examine how Adele's Course 10 criteria differ from Theresa's Course 10 criteria, discussing what that shows about expert-reader expertise.

Criteria overall. As I analyzed Adele's original and redo placement session think-aloud protocols, I focused on unproblematic Course 10 students—those 26 participating students who were placed into Course 10 in both sessions and who were easily placed without being designated as tweeners or flagged as old Course 11A students. As I noted earlier, I counted how many students Adele discussed each criterion on and if that discussion was positive, negative, or mixed. Table 5 details the unproblematic Course 10 students' totals as discussed by Adele for each of the reading and writing criteria.

Table 5

Total Number of Unproblematic Course 10 Students Discussed by Adele for Each Placement Criterion

# of Students	% of Students	Name of Criterion
22/26	84.62%	Reading issues/quality
18/26	69.23%	Assignment task
17/26	65.38%	Essay/ paragraphing knowledge
16/26	61.54%	Working with text
14/26	53.85%	Grammar
9/26	34.62%	Use of required texts
8/26	30.77%	Length/amount of writing
7/26	26.92%	Language
4/26	15.38%	Thinking
3/26	11.54%	Framing/moves
2/26	7.69%	Thesis/project
2/26	7.69%	Print conventions
0/26	0%	Academic rhetorical conventions
0/26	0%	Writing process

Negative rather than positive comments. As Adele discussed the unproblematic Course 10 students, the overwhelming majority of her comments were negative or mixed overall. Of the 26 unproblematic students, only three students had more than one positive criterion noted. The other 23 students broke down to the following: seven students with only negative criteria noted, nine students with one mixed criterion and the rest negative, four students with more than one mixed criteria and the rest negative, and three students with only one positive criterion.

Important criteria for Course 10. The total number of students commented on for each criterion gives a sense of the criteria most important for Course 10. Clearly, reading issues and quality, assignment task, essay and paragraphing knowledge, and working with text are still important factors for Course 10 placement, as they are for Course 11 and placement overall. Additionally though, the criterion of grammar increased in importance for Course 10 placement while the importance of framing and moves fell both in comparison to Course 11 placement and placement overall. While the overall count can help determine those criteria that matter for Course 10 placement, since the four major criteria that were important for placement overall reappear here, we have to look a little deeper into the qualitative end of the data to see (a) what aspects of these particular criteria matter for Course 10 placement; (b) how the breakdown of positive, mixed, and negative comments plays out with specific criteria; and (c) what this all reveals about the relationships among the Course 10 criteria involved in determining placement.

Reading issues and quality. Just as it was with unproblematic Course 11 placement students, the criterion of reading was commented on the most out of all the

criteria for unproblematic Course 10 students. However, unlike Course 11 students, Course 10 students more frequently exhibit concerning or serious misreading with placement readers noting overall misreading at a higher rate. Of the 22 students commented on for reading issues and quality by Adele, 11 received fully negative assessments and were described as “misreading,” “making things up,” or turning the text into “clichés.” For four of these 11 students, additional concern was noted with Adele saying the student had “really bad reading problems” (Redo Session 2, 1010-01-05) or even comments like “it’s like they haven’t read the text at all” (Redo Session 2, 1010-04-01).

While there were a good number of students with mixed assessments on reading, (10 out of 22) and even one student with a positive assessment, these occurrences were far less the norm for Course 10 students. Furthermore, a close examination of what Adele says on some of these students shows that even though they received some comments on reading that were positive in terms of reading quality, the kind of things they were noted as doing well, when coupled with what they were doing poorly, shows that the assessment was really far more negative overall. Two of the 10 students given mixed comments, students 1010-07-04 and 1010-06-11, did not produce an essay that contained any content on Graff’s text. While Adele did say that these students seemed to have understood the text on Pluto, a positive comment, she offered as well the negative reading comment that these students likely avoided Graff’s text because they didn’t understand it at all. So, despite receiving a mixed set of comments, these students actually showed the more serious reading concern of avoiding one text altogether. One additional student of the 10 students with mixed assessments on reading, student 1010-06-05, also had a

positive comment made on his first sentence where Adele noted a “strong opening” that shows he may “understand,” but since the remainder of his writing was just copied from the text, Adele wasn’t sure he really understood at all.

Interestingly, an analysis of these students all show how Adele connected reading to other criteria as she thought about a student’s performance. She considered if the student had used the required texts, as with 1010-07-04 and 1010-06-11, and what that said about their ability to read, if they left out a text completely. And, she thought about what text boundary issues like copying, a part of the criteria of working with the text, mean for students’ understanding of the reading, as with student 1010-06-05. Adele’s positive comments on one student as well show that in addition to looking at the relationship of reading issues to other factors, Adele weighed issues with one criterion against issues with another: for student 1010-04-03, Adele said, about their reading, that “they actually do understand a lot” from the texts, but she noted “the English and the sentence structure is so bad” that “there’s just too much to be done on their writing” (Redo Session 3).

Assignment task. On the criterion of the assignment task, which includes both comments on a student’s understanding of the assignment as well as his or her performance of it, Course 10 students received primarily negative comments with some mixed. Of the 18 students commented on for assignment task issues by Adele, 12 had negative comments, five had mixed, and zero students received fully positive comments. Furthermore, an examination of the negative students showed a good number of these students, eight out of the 12, were described as “switching” the task because they did not understand what they were asked to do. For example, on student 1010-06-11, Adele said

the student writes “all about Pluto” and she added “they have totally misread the assignment” (Redo Session 2). Three out of the 12 students with negative comments on the criterion of assignment task were described as writing about each text separately with no attempt to connect them, as with student 1010-02-01 who talked about Graff and then Pluto, but Adele noted “there’s no connection” made between them (Redo Session 2).

In the transcripts with those students who received a mix of some negative and some positive comments, an analysis revealed students fell into one of two categories in relation to the assigned task. Either they attempted or tried the task—making a connection between Graff and Pluto’s demotion—but did so without enough writing on that connection to be clear on what the connection was (for example, student in Redo Session 3, 1010-07-01) or the connection they were making was built on a misreading of the texts (for example, student in Redo Session 3, 1010-06-06). Students in this final category—who have a failed connection because of misreading—again revealed the relationship of criteria, in this case, the criteria of meeting the assignment to reading, as these students were not able to do the assignment well because of their misreading.

Essay and paragraphing knowledge. Seventeen students were commented on regarding essay and paragraph knowledge. Five of these students received positive assessments, and examining the think-alouds here shows that these Course 10 students were being praised primarily for showing knowledge of paragraphing, as four out of five students received a comment simply saying that the student “has paragraphs” (for example, 1010-01-04 in Redo Session 1). Only one of these five positive students was described by Adele as having “four paragraphs and an essay,” and that was student 1010-01-01 (Redo Session 1).

Twelve students received negative comments and the vast majority of these students, 10 of the 12, had think-aloud comments made by Adele that described them as having “no paragraphing” (Redo Session 3, 1010-06-06) or “one big paragraph” (Redo Session 1, 1010-06-04). Of the other two negatively commented on students, one student had her response described as simply “not an essay” (Redo Session 3, 1010-07-01) while the other student, 1010-01-02, wrote “two separate essays” which Adele says is “a no brainer” signaling “that’s a [Course 10]” (Redo Session 2). When Adele notes that the student’s writing, even with more than one paragraph, separates the two responses, as she does with 1010-01-02, I understood her comment as expressing a negative since the students did not understand that an essay makes one focused point across paragraphs, and since the students did not conceive of their response as separate paragraphs of the same response but rather separate responses altogether.

Framing and moves commented on less frequently. One of the key criteria for Course 11, as I discussed earlier, was the student’s sense of framing and moves, basically the sense that the reader got about the student’s planning and purpose in various paragraphs. Given the observations I noted above about Course 10 students lack of paragraphing and essay knowledge, it is not surprising that this criterion is less frequently observed. However, it is possible that a student’s lack of framing and moves is a powerful signal. As I have hypothesized earlier, the criterion of framing and moves appears to be higher on the continuum of the development of full preparedness for college level work. Students in Course 11, who had paragraphs and essays, also frequently had a purpose and point for their paragraphs and some clear planning of how the moves work together. That was not the case for Course 10 students. For Course 11,

57.69% of students had comments on their framing and moves, and these were overwhelmingly positive, but only 15.38% of Course 10 students had comments on framing and moves, and all of these were negative, discussing points and content that seem to be “thrown in” (Adele, Original Session 5, 1010-07-01) or sequenced only as a retelling or “summary” (Adele, Redo Session 1, 1010-04-04).

Working with the text. The majority of students (11 out of 16) discussed by Adele in regards to how they worked with the text were described as not understanding how to use the text in some way, perhaps not knowing how much work was needed with the text (Redo Session 2, 1010-04-09) or not being able to work the two texts together (Original Session 5, 1010-06-11). It seems on working with the text, Course 10 students show some similarity to Course 11 students, as these issues are related to Course 11 students’ issues with “unpacking” or “close reading.” However, Adele’s comments characterize Course 10 students as slightly different from their Course 11 counterparts.

Course 10 students not only struggle to do enough work with the text to show their reading and connection—unpacking—some of Adele’s comments indicate it is more likely that Course 10 students don’t know how to work the two texts together in writing to figure out a reading or connection. For example, Adele wrote about student 1010-06-04, “They don’t do the work using Graff to read the Pluto paper...They do have some reading of Graff’s text. They just don’t know how to use it” (Redo Session 1). Another key difference between Course 10 students and their Course 11 counterparts has to do with the subcategory of copying. While none of the Course 11 students had issues with copying, four of the 16 Course 10 students commented on for this criterion were described as “copying” or “lifting” whole passages or partial sentences from the texts

without acknowledging these words were taken from the text, for example, student 1010-06-05 (Original Session 6).

Grammar. The percentage of Course 10 students commented on for the criterion of grammar, 53.85%, was significantly larger than the percentage for Course 11 students at 7.69%. This increase suggests that grammar, while not at the top of the list of criteria noted overall in Course 10 placement, is a significant factor at times for determining Course 10 placement. If we return to the two Course 11 students who were commented on for grammar, 1011-05-03 in Redo Session 1 and 1011-09-02 in Original Session 7, we see Theresa's comments were on commas and ESL idiom interference respectively. These two comments were the only two comments made on grammar for Course 11 students, and both were comments on relatively sophisticated grammatical issues.

Course 10 students also receive a majority of negative comments on grammar: 14 out of 26 students had observations made about their grammar, and 13 of these 14 students had negative comments while one received a mix of comments in the redo session but only negative ones in the original session. However, Course 10 students had comments that described more serious grammatical issues than Course 11 students. Adele noted issues with fragments, spelling, sentence structure, subject/verb agreement, and the dropping of tense and number word endings. The implication on a number of these students was that the errors were "serious" (Original Session 5, 1010-07-04) or "interfered with meaning" (Redo Session 2, 1010-04-09), and for the majority, or eight of the 14, the errors were of a type consistent with ESL language issues. This ESL interference caused Adele to say, as she did with student 1010-04-09, that "they would

definitely benefit from being in [Course 10]” to develop their language more (Redo Session 2).

Students with both original and redo session data. There were eight students in this group of unproblematic Course 10 students that participated in both phases of study, and therefore we have think-aloud data on their essays from both the original and redo placement sessions. Looking more closely at these eight students, I investigate the ways Adele’s comments on these students were consistent with each other and how they changed across readings. Six of the eight unproblematic Course 10 students with data from both placement sessions had identical assessments made on two to six placement criteria with the only difference being the addition of one to three criteria included in the original reading not noted in the redo session. The agreed on criteria for these students—1010-01-06, 1010-02-01, 1010-06-02, 1010-06-05, 1010-06-11, and 1010-07-04—always contained at least one of the major Course 10 criteria.

One of the eight students, 1010-06-10, had the same seven criteria discussed, but while five of these criteria noted the same identical assessment, two of these criteria (work with text and grammar) switched from fully negative to mixed assessments between the original and redo sessions. The final student, 1010-07-01, shows the most changes. Two criteria were mentioned in both readings, working with text and assignment task, although the assignment task comments switched from fully negative in the original reading to mixed in the redo session. Additionally though, during the original session reading of this student’s essay, Adele noted four other criteria not noted in the redo session, and in the redo session, she noted one criterion not noted in the original session.

What these numbers demonstrate is a high level of consistency across reading sessions. Adele discussed a good number of the same criteria on the same students across readings. Interestingly, these numbers suggest that most of the differences between Adele's readings can be accounted for by her simply saying more during the original placement sessions than she did during the redo sessions. For six of the eight students, the only difference in reading was additional criteria mentioned in the original placement session. Adele saying more during the original session makes sense given that she knew the redo session was a blind rereading of essays that had already been placed, and the fact that the decisions she was making would not have real consequences could have subconsciously affected Adele, making her less worried about making a decision without noting as many pieces of evidence. The change in the assessment of the other two students—where particular criteria that were given a fully negative assessment were now discussed with both negative and positive assessments—is not very worrying either. Since Adele already noted negative assessments on other major criteria and still noted the same negative aspects on these changed criteria, and since the now positive observations did not cause her to change her decision, it is likely that the amount of negative assessment overall outweighed the positive aspects she now saw in the second reading.

Interesting observations on expertise. Smith (1992/2009, 1993) found that expert-readers, those with recent course teaching experience, were more reliable readers when asked to decide placement into or out of their course of expertise. Given that Cardinal College uses Smith's expert-reader model, it is important to look at the differences between the reading and assessment of the expert-reader for a course and the other reader. With that in mind, I turn to compare Adele's Course 10 criteria with the

criteria Theresa noted as she decided to push these students down from her course of expertise, Course 11. Table 6 details the unproblematic Course 10 students' totals as discussed by Theresa for each of the reading and writing criteria.

Table 6

Total Number of Unproblematic Course 10 Students Discussed by Theresa for Each Placement Criterion

# of Students	% of Students	Name of Criterion
21/26	80.77%	Reading issues/quality
19/26	73.08%	Assignment task
12/26	46.15%	Essay/ paragraphing knowledge
10/26	38.46%	Working with text
9/26	34.62%	Use of required texts
6/26	23.08%	Grammar
6/26	23.08%	Length/amount of writing
5/26	19.23%	Academic rhetorical conventions
4/26	15.38%	Language
4/26	15.38%	Thinking
4/26	15.38%	Framing/moves
3/26	11.54%	Print conventions
0/26	0%	Thesis/project
0/26	0%	Writing process

There are several interesting observations to make on the similarities and differences between Theresa's criteria for pushing these students down and Adele's criteria for placing these unproblematic students into her course. Figure 1 compares Adele and Theresa's Course 10 criteria, displaying the percentage of students commented on for each criterion. There are striking similarities between Adele's criteria for Course 10 and Theresa's. Both Adele and Theresa commented on students most frequently for the same top four criteria: reading quality, assignment task, essay/paragraphing knowledge, and work with texts, but there were also some differences in the percentage of students commented for these issues. While a student's reading quality was a good

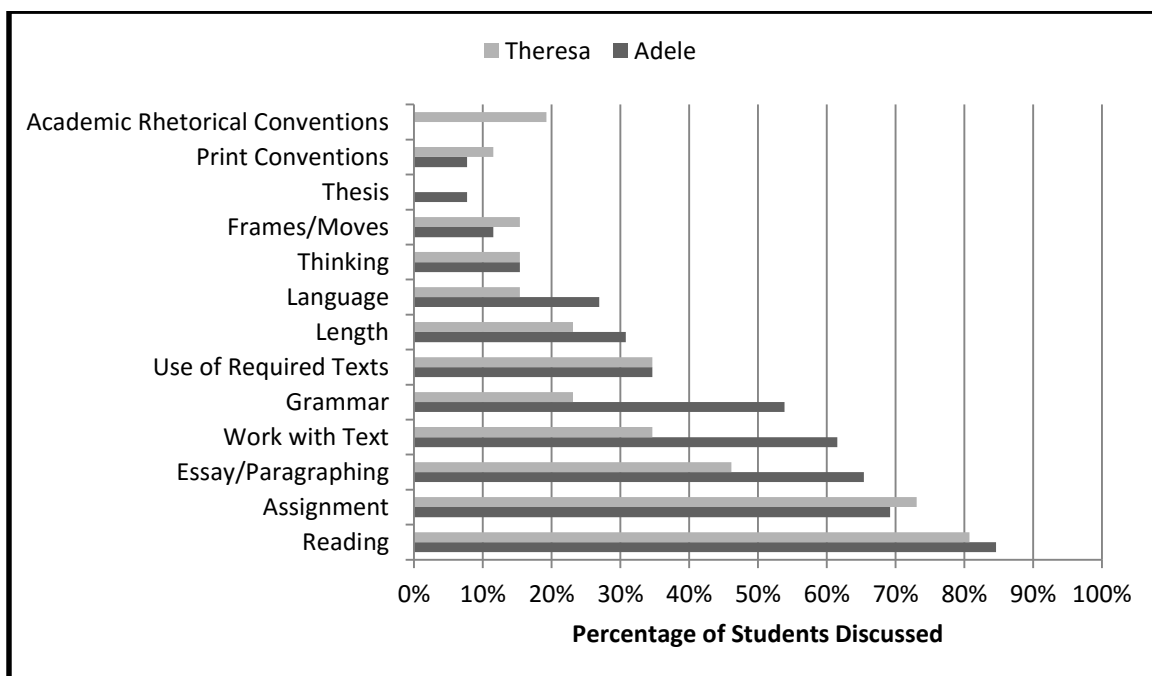


Figure 1. Comparison of readers' Course 10 criteria.

deal more important than their understanding and meeting of the assignment task for Adele, these criteria were closer in terms of the percentage of students discussed on the issue for Theresa. Furthermore, while both readers commented on essay and paragraphing knowledge, the third most frequently overall, there was a fair difference in how frequently; Theresa commented on this issue for 46% of the students to Adele's 65% of students.

The most striking differences between Adele and Theresa here have to do with grammar and language comments. Comments on grammar by Adele were significantly higher than Theresa, and this issue was the fifth most frequently commented on criteria for Adele. For Theresa, the issue of grammar dropped to sixth overall, and slightly more than thirty percent fewer students were commented on by Theresa for grammar than Adele. While there were differences here between Adele and Theresa's frequency of

comments on academic rhetorical conventions and thesis—only Theresa commented on rhetorical conventions and only Adele commented on thesis statements—since so few students received comments on these issues overall, these criteria do not seem as important.

In short, there were some notable differences between these readers that could potentially be related to their different course expertise. To investigate this further, let's compare Theresa's Course 11 criteria, the course for which she was reading as the expert, to her Course 10 criteria, the course for which she was not. Figure 2 compares Theresa's Course 10 and Course 11 criteria, and it reveals how Theresa's reading of Course 10 essays differs from her reading of Course 11 essays.

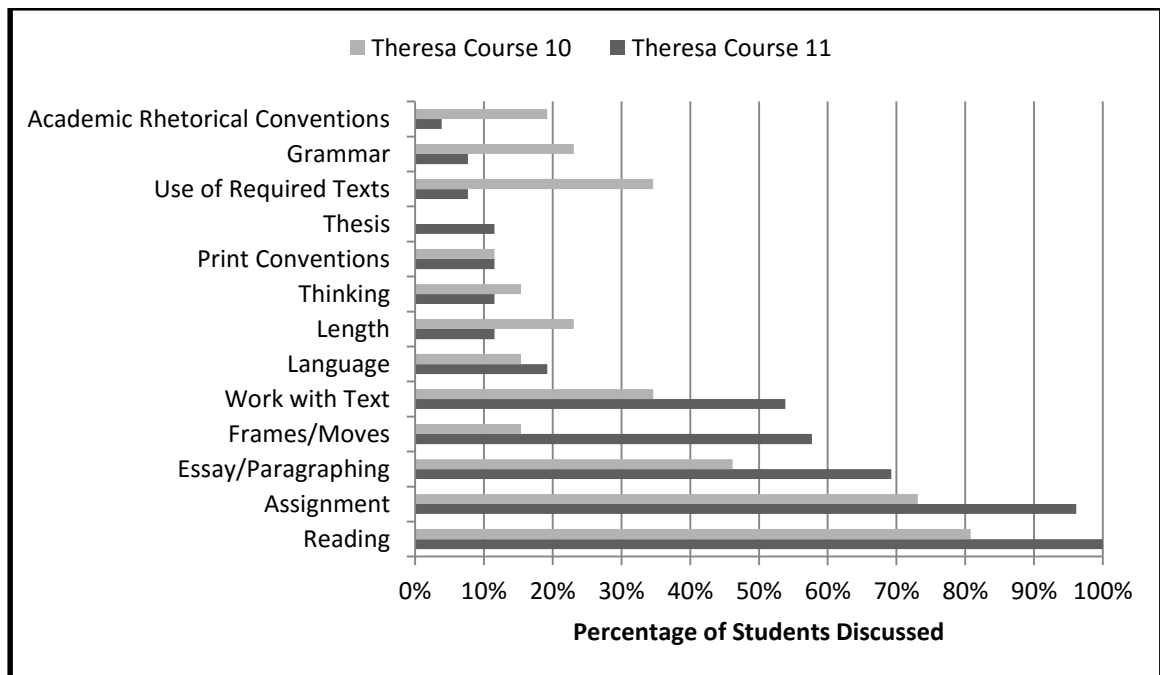


Figure 2. Comparison of Theresa's Course 10 and Course 11 criteria.

While the same three criteria occupy the top spots for both readers—reading, assignment task, and essay/paragraphing knowledge—all three drop in frequency overall

when Theresa reads Course 10 essays. In addition, there were stark differences in the discussion of students' frames/moves, grammar, and use of the required texts. What is shown by comparing Theresa as a reader of essays placed into the two different courses is how even though her and Adele have differences, they are much more in tune when reading unproblematic Course 10 students than it seems on the surface. To pull one illustrative example, while Adele and Theresa differed significantly in how frequently they discussed grammar on Course 10 students' essays, Theresa as a Course 10 reader also differed significantly from herself as a reader of Course 11 essays. It seems that as a Course 10 reader, Theresa moves closer to the expert Course 10 reader.

We will continue to discuss the differences between these readers on expertise as we go forward to consider the criteria of the more problematic students, but I believe that there is some influence here regarding expertise, as Smith (1992/2009, 1993) theorized. I also think some of Haswell and Wyche-Smith's (1994) findings hold true here as well. Building on Smith (1992/2009, 1993), Haswell and Wyche-Smith (1994) suggested that there are some student essays on which placement decisions can more easily be made—those students who better fit (or in this case, do not fit) the course—and even less expert readers will not experience significant disagreements placing these essays. What we are looking at here is precisely this bunch of unproblematic essays—students who neatly fit Course 10 and who, thus, do not fit Course 11. Not one of these students, in two separate reading sessions, was hesitated over or flagged in any way, and the same decision was made both times. It remains to be seen what happens when we examine those students who do not fit neatly into the available courses.

Edge Placement Criteria

Turning to problematic students—I discuss the edge criteria, or the criteria that represent the border between the two Cardinal College courses, as determined by an analysis of all original and redo session think-aloud protocols on problematic Course 11 student essays and problematic Course 10 student essays. Remember, I made the decision to separate the participating placement students into unproblematic and problematic groups because in two separate reading sessions, some students were cleanly and consistently placed while others were not. I felt the data on those students who were consistently placed would make good representatives of the criteria for their corresponding courses. However, I believed the problematic students—students who were not easily placed, as they were designated as tweeners or flagged as old Course 11A students in one or both sessions, and students who received a different placement decision in the redo session than their actual course placement—could be examined to better understand which criteria or combinations of criteria in student essays present readers with placement difficulty.

Of the total 108 participating students, there were 26 problematic Course 11 students and 30 problematic Course 10 students. In this group, there were a total of 45 students flagged as old Course 11A (either in one or both of the original and redo sessions), 13 tweeners (five from the original sessions and eight from the redo sessions), and 18 students who had a different placement decision made about their essay in the redo session. See Appendix S for a list of all 56 problematic students and a breakdown of their problematic designations, including a number of students with multiple problematic designations.

I begin by examining the problematic Course 11 students, showing overall counts in criteria and trends in positive versus negative comments, as well as the key criteria involved with problematic placement decisions, before doing the same for problematic Course 10 students. Since many of these students' essays were read by both readers, and since there were problems with the placement decisions on these essays, I make use of both readers' criteria, but I provide the expert-reader's criteria first and give primary weight to it, contrasting it where necessary with the criteria of the other reader. As I detail both the problematic Course 11 and problematic Course 10 criteria, I discuss similarities and differences between these problematic students and unproblematic students of the same course, making sure to contrast this discussion with the comments on placement made earlier, and I note and discuss important students on whom I have both original and redo placement data. Once finished, I discuss similarities and differences between the two groups of problematic students themselves.

Problematic Course 11 criteria for Theresa. As I analyzed the original and redo placement session think-aloud protocols, I first focused on the total 26 problematic Course 11 students which were all read by Theresa. As I noted earlier, I counted how many students each reader discussed each criterion on and if that discussion was positive, negative, or mixed. Table 7 details the problematic Course 11 students' totals for each of the reading and writing criteria as discussed by Theresa, the expert-reader.

More mixing of positive and negative comments for Theresa. Ten of the 26 problematic Course 11 students discussed by Theresa received no positive comments at all on any criteria. Thirteen of the 26 students had at least one criteria assessed positively, and some had more than one, but the number of criteria discussed in negative or mixed

Table 7

Total Number of Problematic Course 11 Students Discussed by Theresa for Each Placement Criterion

# of Students	% of Students	Name of Criterion
26/26	100%	Reading issues/quality
20/26	76.92%	Assignment task
20/26	76.92%	Essay/ paragraphing knowledge
19/26	73.08%	Working with text
12/26	46.15%	Framing/moves
11/26	42.31%	Length/amount of writing
7/26	26.92%	Academic rhetorical conventions
5/26	19.23%	Language
5/26	19.23%	Thinking
4/26	15.38%	Use of required texts
3/26	11.54%	Print conventions
3/26	11.54%	Grammar
0/26	0%	Thesis/project
0/26	0%	Writing process

terms outnumbered those criteria discussed positively for these students. The remaining three students had a higher number of positive criteria noted than negative or mixed.

When Theresa discussed unproblematic Course 11 students, recall, the overwhelming majority of her comments were positive on some criteria or mixed on others. But, with problematic Course 11 students, Theresa's comments were now more mixed or negative overall. Furthermore, on criteria that were more positively discussed for unproblematic students, for example assignment task, Theresa's comments on problematic students were now more mixed. And on those criteria that were already mixed for unproblematic Course 11 students, for example work with text, Theresa's comments on problematic students were now more negative. In short, problematic Course 11 students do not have full competency on key criteria in comparison to Course 11 students.

Problematic Course 11 criteria for Adele. As I analyzed the original and redo placement session think-aloud protocols, I also focused on eight problematic Course 11 students read by Adele. Adele read only eight of these students as that was the number of students who were placed in Course 11 but were tweened in the original and redo sessions or who were pushed down in the redo session and had a different placement decision made on their essay. As I noted earlier, I counted how many students each reader discussed each criterion on and if that discussion was positive, negative, or mixed. Table 8 details the problematic Course 11 students' totals for each of the reading and writing criteria as discussed by Adele. Even though Adele is not the expert-reader for the course, her criteria are still important to consider since three of these students had their course placements changed between the two reading sessions.

Table 8

Total Number of Problematic Course 11 Students Discussed by Adele for Each Placement Criterion

# of Students	% of Students	Name of Criterion
8/8	100%	Reading issues/quality
7/8	87.5%	Assignment task
7/8	87.5%	Essay/ paragraphing knowledge
7/8	87.5%	Working with text
4/8	50%	Grammar
3/8	37.5%	Framing/moves
3/8	37.5%	Thinking
2/8	25%	Language
1/8	12.5%	Use of required texts
0/8	0%	Length/amount of writing
0/8	0%	Print conventions
0/8	0%	Thesis/project
0/8	0%	Academic rhetorical conventions
0/8	0%	Writing process

More mixing of positive and negative comments for Adele. Like Adele's comments on unproblematic Course 10 students, there were a smaller number of fully positive comments with no fully positive comments for four of the eight problematic Course 11 students discussed by Adele here. But, just as with Theresa's discussion of problematic Course 11 students, Adele's comments on these students showed more of a mix rather than the primarily negative comments she offered on Course 10 students. On five of the nine criteria she discussed, the number of students receiving mixed or positive comments outnumbered those with negative. Furthermore, these were criteria on which regular Course 10 students would receive negative assessments by Adele, for example, reading, assignment task, and essay and paragraphing knowledge. Adele's positive to negative comment breakdown suggests that these students have more competency on key criteria in comparison to her typical Course 10 students just as Theresa's comments suggest they have less competency than typical Course 11 students.

Interesting differences between Theresa and Adele as readers of problematic Course 11 essays. Since Theresa and Adele were reading with different recent course teaching experience (Smith 1992/2009), it is not surprising that their criteria were different. Figure 3 compares Theresa and Adele's problematic Course 11 criteria, displaying the percentage of students commented on for each criterion.

As I've already indicated, the same four major criteria are the most frequently commented on criteria in terms of overall students—reading, assignment task, essay and paragraphing knowledge, and work with text. However, Adele discussed these key criteria more frequently than Theresa with this group of students. Part of the reason for this difference may be that Adele commented on fewer criteria overall, only discussing

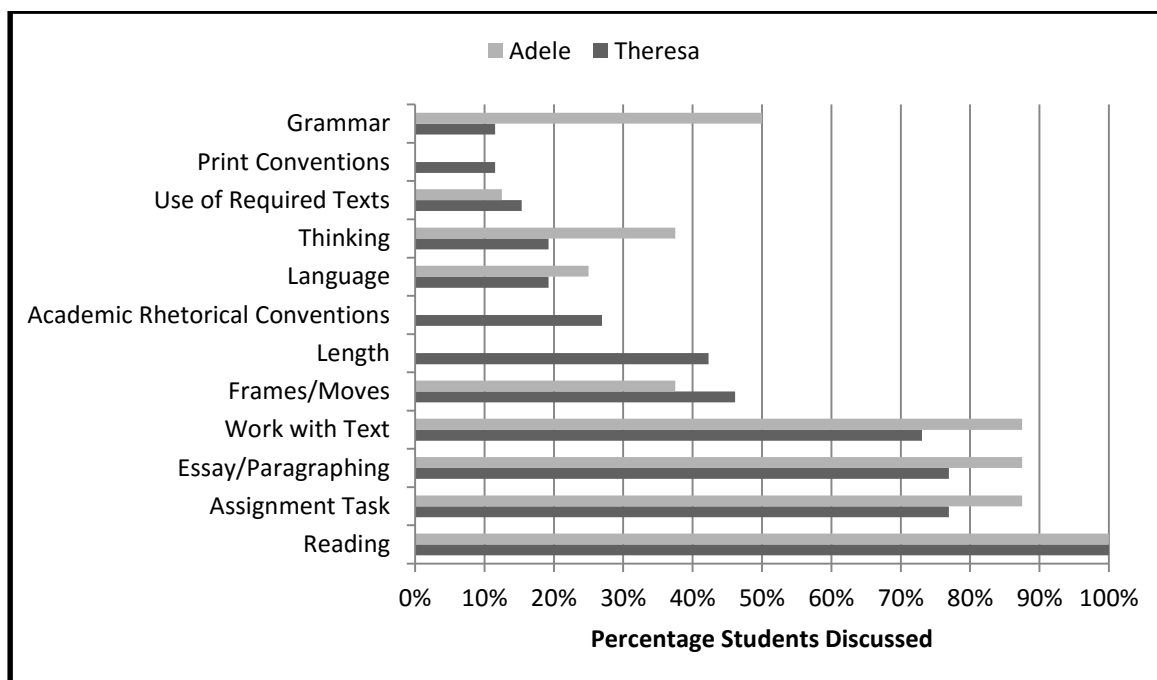


Figure 3. Comparison of readers' problematic Course 11 criteria.

nine criteria to Theresa's twelve. Adele did not discuss length, academic rhetorical conventions, or print conventions with any of these problematic Course 11 students. Finally, Adele discussed grammar and thinking more frequently on problematic Course 11 students' essays than Theresa did, but since grammar was noted more frequently on Adele's Course 10 students, it is not surprising to see that she still talked about that criterion here.

Key criteria that present problems for Course 11 placement. Looking at the total number of students commented on for each criterion is a fairly easy way to determine the criteria most important for problematic Course 11. But we find the same four key criteria—reading, assignment task, essay and paragraphing knowledge, and work with text—at the top of both Theresa and Adele's lists, and these have been consistently important criteria for all of placement. After the top four criteria in Theresa's

list, the next most frequently commented on criteria was twenty-seven percentage points lower than these top criteria. Yet, the next two criteria of framing/moves and length seem more important for these problematic students, despite this drop, since they were still discussed on over 40% of these students' essays. Additionally, Adele discussed grammar on 50% of students' essays, and on both framing/moves and thinking, she discussed 37.5% of students' essays. While looking at the totals is one important way to determine the criteria that matter in placement, we have to look a little deeper into the qualitative end of the data to see what characteristics of these particular criteria—including their positive, negative, or mixed determination—present difficulty in determining Course 11 placement.

Reading issues and quality. Two of the 26 problematic Course 11 students commented on by Theresa for reading were given positive assessments. The two students positively assessed on reading, 1011-07-01 and 1011-02-08, both had comments by Theresa that praised their reading for doing something uncharacteristic of many other readers. The first student did not misread Graff's use of sports as the topic of his text (Redo Session 1), and the other student saw that debate is the key to understanding Graff's text (Redo Session 1). While these were certainly positive characteristics, Theresa said little else on their reading overall as she turned to look at other issues with each student. The same was true of the two students given negative assessments by Theresa. Student 1011-08-01 was described as misreading Graff, but not much else was said once Theresa turned her concern to other aspects of the student's response (Redo Session 1), and student 1011-02-01 "really struggle[d] with Graff" according to Theresa (Redo Session 1).

The far more interesting and important group for seeing what characteristics cause Course 11 placement difficulty in regards to reading are the 22 students given mixed assessments out of the total 26. Looking more closely at their transcripts reveals what characteristics of reading these students exhibited. Eight of the 22 students had comments that suggested they have some misreading of and some good ideas on Graff's text, but as they progress to trying to read the Pluto text with Graff's ideas as a lens, their reading actually improves. For example, with student 1011-05-04 in Redo Session 2, Theresa indicated that the student's reading of Graff is off on the idea of "the world being more interested in sports than school" and she said, "Okay, but [the student] does see the Pluto thing as examples of rivalries and competition in figuring that out." Since four of these eight students were tweeners in either the original or redo session, and one had a different decision made about his placement in the redo session, placing him in Course 10, it seems this kind of reading is concerning.

One of the comments Theresa makes while discussing student 1011-09-06 is also revealing. Examining 1011-09-06, a twener in the redo session that both Theresa and Adele indicated has this type of reading that improved toward the end, shows a little about why it is hard to place such students and why the kind of reading that only begins to emerge at the end of the writing is concerning. About this student, Theresa said, "If this paper were revised and put through some kind of process, they would probably be alright" (Redo Session 2). Here, she suggests that these students may be able to take the reading they arrived at and revise, as they would be asked to do in the course, but the implication of the word "probably" is that it is also possible that such students might also struggle to use revision that way.

Five of the 22 students who were given mixed comments on reading had the opposite problem. These students seemed to do well with Graff's text, or at least partially so, but when they turned to use that reading to look at the situation of Pluto's demotion, their reading started to fall apart a little. For example, with student 1011-02-06 in Redo Session 1, Theresa praised the student's earlier explanation of Graff focusing on sports as an example or way Graff illustrates his point, but as Theresa read the student's comments on Pluto, which suggested the demotion was just to keep the solar system from "growing too large," she said, "Oh! He was on the way. He almost had the horse in the barn, and then the horse took off!" Since none of these students were flagged as tweeners or had different decisions made about their placement—i.e., they were only flagged as old Course 11A—I believe this characteristic, while concerning, is not as serious a concern. In fact, one of the students, 1011-07-02 was described as having a more mixed reading of Graff's text, a less positive description than the other four students, but also was described, like the other four students, as falling apart when trying to use those ideas to read the Pluto situation. He wasn't pushed down in either reading or given a different placement decision either, which suggests that even with some additional reading concerns, these students may look more like their unproblematic peers on the criteria of reading than others in this group.

Another four of the 22 students were simply discussed as having mixed reading, some understanding and some misunderstanding, with one of these students having been a tweener in the original placement sessions. Given that there was only one student here that was a tweener in the original sessions and the rest were all old Course 11A students, it is not clear if this characteristic of reading is seen as more or less serious than the other

characteristics. And sadly, this student, 1011-08-07, was not participating during the original placement sessions, so there was no think-aloud record to show the reasoning Theresa gave for pushing him down to Course 10, nor Adele's reasoning for pushing him back up to see if the original assessment aligns with the assessment given in the redo session or exposes some other characteristics instead.

Three of the 22 students given mixed assessments on reading by Theresa reveal a relationship between reading and working with the text. It was suggested in the discussion of these students that Theresa *guessed* they understood, but she was unsure because of how little work had been done with the text. For example, with student 1011-09-08, Theresa "think[s] that this person gets it," meaning the student understands the text, but the student "hasn't learned how to close read" (Original Session 6), an assessment that Adele expressed in her reading of this student's response as well. One of these three students was unique in that Theresa's assessment also included comments that there was "some fairly major misreading here," but she stressed, "I don't know that it's misreading. It may be that all he needs to be told is to pay attention to the text" and that he should "unpack" (Original Session 6, 1011-09-01). In short, she was unsure if the student had problems or just needed to work more with the text and would solve the seeming misreading. One additional student from earlier, 1011-08-01, who Theresa pushed down in the redo session largely because of a negative assessment characterized as misreading, may fit into this category as well. When Adele read 1011-08-01's essay in the redo session, she ended up pushing him back up, concluding that they "just don't do the work" with the text, so can go back up (Adele, Redo Session 1).

The remaining two students commented on for mixed reading each had problems of their own particular variety, and I am not surprised, given how unique their cases are, that they both had a different decision made about their essay in the redo sessions. Student 1011-02-03 was the only student in this group whose reading was described in mixed terms—some good ideas and some negative—but who also didn't mention Graff's text at all. This absence of Graff means that the student had a mixed reading of the simpler text on Pluto, which would definitely be concerning enough, on its own, to push down. Yet, the student was still placed into Course 11 in the original session. Why? The student was participating in both the original placement session and redo, so we have think-aloud data on both. In the original session, Theresa noted that there was no Graff and speculated that perhaps "Graff was inconvenient" so the student avoided him, as other students do. But as she reread, she noticed "there's a lot of places where she could have talked about Graff," which caused Theresa to decide this was "an assignment issue and not a [reading] ability issue" where the student "needs to be told" they can't ignore the assignment (Original Session 6, 1011-02-03).

In contrast, in the redo session record on student 1011-02-03, Theresa again noticed the absence of Graff, and she said, "This is the mystical part of placement. What does [the lack of Graff] mean? It means that he can't read it? Maybe?" She looked through the piece some more, but ultimately she said, "When somebody doesn't do half of the placement assignment, you can't put them in the top course" (Redo Session 2). It is interesting to note as well that while reading this student's essay, what Theresa said expresses a difficulty deciding during both placement sessions. In the original, she described the student as having both Course 11 and Course 10 characteristics and went

back and forth on deciding. In the redo session, she said that the student's piece overall "looks very much like early Course 11 work" before saying that she will push the essay down (Redo Session 2). As Adele read student 1011-02-03's response in the redo session, she saw the student's reading as largely "words of advice" (Redo Session 2). Although the student did have quotes, Adele saw these as filler that never get explained.

Student 1011-02-07 also had characteristics that were strikingly unique in that rather than turning to the text to offer a reading, the student made up a hypothetical situation involving camps of people waging a debate over if peanut butter is better than jelly—what the student called "the great, grand PB&J debate"—that the student likened to the debate over Pluto (Redo Session 3). As Theresa reacted to this student's essay, she described this type of reading, much like the students who need more unpacking or close reading of the text directly, as a lack of knowing "how to ground things" in the text (Redo Session 3). While we do not have the original session record to look at, we know Theresa decided to place the student into Course 11 originally. In the redo session, however, she went back and forth with the decision, first saying she thinks it "is in," but then saying she "wants to see what Adele does with it," suggesting some doubt in her sense of it, and she pushed it down. As Adele discussed it, she placed it into Course 10 because the peanut butter and jelly debate "is like not a reading" and the student struggled with essay structure as well, having "one big paragraph" (Redo Session 3).

To sum up, the issue of reading is key to the placement of problematic Course 11 students, but for many of these students, particularly those who display performances that are uncharacteristic of the course, we see that assessing their reading is difficult because of a number of issues. First, reading in this context means not just understanding one

text's points, but it means using those points to read another text—to see and understand the second text in the terms of the first. The result is that students who struggle with this more complex kind of reading cause concern for Theresa. Second, the criterion of reading, in this context, is related to characteristics of the writing—the student must work with the text to show they truly understand, and the student must discuss both required texts to show they fully understand and can use the ideas in one text to read another. Without these other writing attributes, Theresa doubts the strength of the student's reading. Third, Theresa links issues with reading directly to the work of the course, as she seems to weigh what will happen with some students when they are asked to revise and for them, that revision means reworking their reading. Will they struggle to do that, and if so, should they be placed in a course that will offer them the additional instructional time of six credits for guidance and scaffolding?

Assignment task. The discussion of reading above suggests that reading comments address the two levels of reading required by the assignment, understanding Graff and being able to read the Pluto text with Graff's ideas in mind. Given these levels of reading, it becomes even clearer than it was when discussing the unproblematic Course 11 students that the assignment task comments are intimately connected to reading comments, a fact revealed further by how the problematic Course 11 students are discussed. Part of a student's success at meeting the assignment is tied up in the limits of their reading. For those students with negative assignment task comments, this is obvious: student 1011-02-03 was missing one text completely and was described as only doing "half the assignment" by Theresa (Redo Session 2) and student 1011-03-02 had two separate essays that offered no connection between the texts, never reading one in

relation to the other. Essentially, these students could not do the assignment without the reading connection.

The overall breakdown of students commented on with mixed assessments by Theresa also bears out the connection of assignment comments to those on reading. Two of the 13 students with mixed comments (1011-03-03 and 1011-08-02) only indirectly connected the texts together or only indirectly worked with one text, making it hard to achieve the analysis of one text through the other. Theresa also shows the link when she noted how poor reading affects the students' ability to do the assignment for 3 of the 13 students with mixed comments. For example, with student 1011-09-14 in Redo Session 1, Theresa said the student "reads Pluto and Graff together, but because she somewhat misreads Graff" her success is affected. For the other eight of 13 students, Theresa's assessment was described primarily in terms of the student figuring out some reading of the two texts together, and as such, these comments barely separate a student's attempt from his or her reading. For example, with student 1011-08-06 in Redo Session 2, Theresa first described the student's reading of Graff. Then, she shows he tried the assignment by saying he "goes on into the Pluto," and finally she noted the success in terms of what he gets to in the connection, which is "he gets into rivalry and different viewpoints." In essence, to do the assignment is to read to a connection. It seems that the large number of mixed assignment comments by Theresa for problematic Course 11 students is directly related to the number of mixed comments on reading.

For Adele, comments on assignment task were also an extension of reading as most of her assignment comments, five out of seven, indicated the student tried to address the assignment because they attempted to name a connection between Graff and the Pluto

situation, just like Theresa's comments showed. For example, Adele noted that student 1011-03-01 has "some misreading in here, but there are also some good things. They also manage to try, like mesh or integrate Graff's text with the Pluto text" (Redo Session 2). And, even on the negative side, the assignment was described in terms of the reading. Adele, just like Theresa, said student 1011-02-03 has no Graff with which to do the assignment and read for a connection. Yet on student 1011-03-03, who Theresa believes was indirectly using the text, saying his discussion of Pluto "might have been inflected some by the idea of hidden intellectualism" from Graff, Adele said the student makes "no connection really to what Graff is talking about" (Redo Session 1).

While both Theresa and Adele presented a more mixed view of these problematic Course 11 students than their unproblematic peers, and while both displayed assignment comments that showed the intimate connection of reading to the assignment, some of the comments here also give us a more specific sense of what it means to strongly address the assignment, taking the addressing of the assignment further than its relationship to reading. As Theresa discussed five students with positive comments on their ability to address the assignment, her comments on several of these students go past just the successful attempt at a connection. And, these comments say more about strongly addressing the assignment than was shown during Theresa's reading of unproblematic Course 11 students. For example, with student 1011-03-06 in Redo Session 1, Theresa said the student writes "What people do not realize is that this is what keeps the real intellectual world going. If someone doesn't step up and disagree, that will leave no room for new discoveries" to which Theresa commented "That's actually pretty profound!" What we see is that the strongest achievement of the assignment goes past making a

connection and having a reading alone. The praise given to student 1011-03-06 is Theresa's excitement that the student saw a connection and used it to comment back to what that means for how our ideas and knowledge change.

Essay and paragraphing. Unlike reading and assignment comments, which were more mixed on problematic Course 11 students than their unproblematic counterparts, Theresa's comments on the problematic Course 11 students' essay and paragraphing knowledge take a more negative turn than unproblematic Course 11 students. Twelve of the 20 students commented on for essay and paragraphing knowledge received negative comments overall, six received positive comments, and two had a mixed assessment. The increase in negative comments on essay and paragraphing could very well be one reason why these students' essays presented placement difficulty or concern, since nine of the 12 students with negative assessments had only one paragraph, aligning them more with Course 10 students than Course 11 ones. Yet a negative assessment on essay and paragraphing knowledge alone obviously wouldn't be enough of a concern to push a student down to Course 10, or the twelve students with negative assessments on this criterion would have had their decisions more easily made.

Examining Theresa's comments on these problematic Course 11 students with negative assessments here reveals that concerns about essay and paragraphing are weighed against a number of other criteria, most importantly reading. For example, with student 1011-02-06 in Redo Session 1, Theresa expressed concern that the student's response was only one paragraph, saying "And, it's only one paragraph. Oh, my dear, Jesus!" but her comments came directly after noting that the student was doing some good reading, albeit with noted problems. In the end, the decision came down to reading

as Theresa claimed the student was “reading as well as my [Course 11s] at the start of the semester.” With other students who write only one paragraph, Theresa weighed the paragraphing with concerns about reading along with those about the students’ assignment issues or need for unpacking. For example, with student 1011-03-06 in Original Session 5, Theresa weighed reading and unpacking with her concern about the one paragraph response. After noting that the student had only a paragraph, she said that “this person really gets the idea” but “doesn’t have any idea how to work with a quotation.”

Theresa’s weighing of essay and paragraphing knowledge against other criteria is consistent with Adele’s treatment of the criterion of essay and paragraphing as well. For example, with student 1011-03-06, also discussed by Theresa above from Original Session 5, Adele embedded a comment about the student’s paragraphing inside a discussion of multiple other criteria, like assignment task, work with text, and reading. She said:

They really try and integrate. They work with the two texts. They really try to use Graff’s text to interpret what’s going on in the Pluto text. And it’s throughout the entire essay. There’s no organization, no paragraphs or whatever, but that’s not a big deal. (Adele, Original Session 5)

The student’s one paragraph response was not seen as enough of a concern to place them in Course 10. For another student though, 1011-02-07 in Redo Session 3, the weighing resulted in the opposite decision. Adele discussed some problems with the student’s reading, and then she said, “This is like one big paragraph too. So [it’s a problem] on all

fronts,” suggesting that her concerns with the student’s reading along with the paragraphing strengthened her sense that the student belonged in her course.

Relationship of essay and paragraphing knowledge to framing and moves.

Unproblematic Course 11 students, as I discussed earlier, were commented on frequently by Theresa regarding their framing and moves. And while the percentage of problematic Course 11 students commented on for this criterion dropped over eleven percentage points from unproblematic Course 11 students, Theresa still commented on framing and moves for 46.15% of these problematic students. In addition, the characteristics of these comments changed from being primarily positive with some mixed on the unproblematic students to nearly a half way split between positive and negative assessments on the problematic students. This change in frequency of negative comments is related to the same change in the number of negative comments provided on essay and paragraphing knowledge.

Just as with unproblematic students, Theresa’s comments on framing and moves coincide with comments on essay and paragraphing: all 12 of the students commented on for framing and moves also received comments on their essay and paragraphing. In addition to the comments on these two criteria coinciding, they frequently line up in the positive or negative quality of the comments. Of the 12 students commented on for both of these criteria, seven had assessments that lined up. For example, with student 1011-02-07 in Redo Session 3, Theresa said the student “needs some help in shaping an essay and framing [their hypothetical PB&J debate] as an example instead of getting lost in the story of the PB & J which he sort of does,” indicating that the student has trouble with essay structure and planning moves. On the positive side, with student 1011-02-03 in

Original Session 6, Theresa said first that the response “looks kinda sorta like an essay” with “four paragraphs.” Then she indicated that in addition to a positive essay assessment, the piece also had moves: “It has first reason, second reason. You know if you look at the opening of the paragraphs.” For the remaining four students with both these criteria commented on, the quality of the comments is different between the criteria. For example, with student 1011-09-01 in Original Session 6, Theresa said, “it’s one paragraph,” but adds later “this kid has some framework.”

Work with the text. For problematic Course 11 students, the most important criteria for determining placement, after the criteria of reading, is how a student works with text. As I’ve already discussed with earlier groups of students, this criterion for Cardinal College has a lot to do with the criterion of reading as comments on a student’s work with the text center on the student’s explanation of what they see and understand from the text. Therefore, it is not surprising that Theresa’s discussion of these two criteria are intimately linked, and it is not surprising to find that work with text was an important criterion for all placement decisions. But, this criterion was especially important for problematic Course 11 students because the way these students work with text, or fail to work with text, could cast doubt on how much understanding they had of what they’ve read.

For unproblematic Course 11 students, 11 out of 14 comments by Theresa on working with the text were positive. But, for problematic Course 11 students, there were no positive comments, and 12 out of 19 students received negative comments. In addition, putting the 19 comments in categories, based on what aspects Theresa’s comments were assessing, reveals the specific issues of concern here. Nine of the 19

students were discussed as needing to offer a “close reading” (1011-05-04, Original Session 7) or “unpack” (1011-05-06, Original Session 6), and most of these students were described by Theresa as having good quotes that just required explanation, as with student 1011-05-04 above, or required more explanation, like student 1011-05-09 (Redo Session 2).

For four of the 19 students, the comments were similar to the call for unpacking, but the wording suggests Theresa was more worried about the student’s reading. Therefore, the student needed to work more directly with the text, providing quotations or specific enough details from the text since their reading was “invisible” and needed to be shown (1011-03-07, Redo Session 2). One additional student of the 19 students was described as having good quotations, but “the wrong conclusion” about the quotations (1011-08-01, Redo Session 1), a description also related to these other students in its concern with the student’s reading. In this case, the student was seen by Theresa as needing, like his counterparts, to focus more on what the text actually said to arrive at his understanding.

If the earlier 14 students’ comments centered on how they worked with quotes or the need for quotes, the next two students of the 19 had comments that discussed summary or paraphrase instead (1011-07-01 and 1011-08-07). For both students, Theresa noted they were relying heavily on summary or paraphrase, the implication being that they relied on these rather than quotations. But both students were described as having good ideas and enough of a sense of what they were trying to explain about the text in these moments of summary or paraphrase that it was clear they understood. For example, with student 1011-07-01 in Redo Session 1, Theresa said the student falls into a summary

“that is not necessarily needed,” but it seems to “have a point in a way” and contains a place where the student is “in a position to turn” their reading back to Graff. The final three students showed the most serious issues, issues with text boundaries, with one using quotations without marking boundaries (1011-03-01), one only using Graff’s text indirectly (never crediting ideas to him or acknowledging his text), and one who had no Graff at all, but who had, at least as interpreted in the original reading of the student’s response by Theresa, moments that could be related to the ideas of Graff (1011-02-03).

Interestingly, with these problematic Course 11 students, Adele’s record becomes extremely valuable for illustrating how important a student’s work with text is for his or her placement decision. The eight students read by Adele were all pushed down by Theresa, seven in the redo session and one in the original session (although this one was not pushed down in the second redo session reading). Five of these students ended up being tweeners, that is Adele pushed them back up. Three of these students were placed into Course 10, in short, having a different decision made about their placement in the redo session than was made in the original. Since seven of these eight students had comments made by Adele on their work with the text, it is fairly easy to compare which students were given a more promising assessment. For the three students who were placed into Course 10—1011-02-03, 1011-02-07, and 1011-03-03—Adele indicated these students did not have quotations or they had “quotations in there, but there’s no reading in there” (1011-02-03, Redo Session 2). Because of this lack of explanation, Adele doubted their reading: “Was [the lack of explanation] intentional, or do they just not get it” (1011-02-03, Redo Session 2). But the other four students she commented on for this criteria were all described as not “sticking” (1011-03-01, Redo Session 2) to the

reading or not doing “enough” with the quotes. And on these students, Adele felt more assured that they had some ideas from the text to build on: “Good thinking going on here, and that thinking is based on what they’ve read” (1011-03-06, Original Session 5).

Length. For Theresa, length was commented on more frequently for problematic Course 11 students than for unproblematic Course 11 students, increasing from 11.54% of the unproblematic students to 42.31% of the problematic ones, with the majority of comments being negative. Since five of the 11 problematic Course 11 students commented on for length also had comments on their responses regarding work with text which claimed that they needed more quotations and explanation, I believe the issue with length was a signal to Theresa that these students may lack the proper development of their reading.

Problematic Course 10 criteria for Adele. As I analyzed the original and redo placement session think-aloud protocols, I next focused on 19 problematic Course 10 students read by Adele out of the total of 30 problematic Course 10 students. The reason Adele did not read all 30 of these essays has to do with the total number of students on which a different decision was made during the redo session and of these, how many were participating in the original placement sessions. Fifteen of the total 30 problematic Course 10 students had a different decision made on their placement, and 13 of these students’ essays were never pushed down during the redo sessions for Adele to read. However, two of the 13 students were participating in the original placement sessions, and therefore, Adele’s original reading can be factored in here. The result is that 11 problematic Course 10 students’ essays were read by Theresa, but not by Adele. Given this large difference, I counted how many students each reader discussed each criterion

on and if that discussion was positive, negative, or mixed. Table 9 details the problematic Course 10 students' totals for Adele, the expert-reader.

Table 9

Total Number of Problematic Course 10 Students Discussed by Adele for Each Placement Criterion

# of Students	% of Students	Name of Criterion
19/19	100%	Reading issues/quality
18/19	94.74%	Assignment task
14/19	73.68%	Working with text
12/19	63.18%	Essay/ paragraphing knowledge
7/19	36.84%	Framing/moves
5/19	26.32%	Grammar
4/19	21.05%	Language
4/19	21.05%	Thinking
3/19	15.79%	Use of required texts
2/19	10.53%	Length/amount of writing
2/19	10.53%	Print conventions
2/19	10.53%	Academic rhetorical conventions
0/26	0%	Thesis/project
0/26	0%	Writing process

More mixing of positive and negative comments for Adele. Adele, as a reader of regular Course 10 students, those without problematic designations, primarily offered negative or mixed comments. Very few unproblematic Course 10 students received more than one fully positive comment on their criteria. But now, for problematic Course 10 students, Adele's comments were more mixed for several of the key criteria, namely reading and the assignment task. For the other criteria, the number of negative comments, while still dominant, decreased, and mixed comments increased with the exception of work with text comments which still remained highly negative.

Problematic Course 10 criteria for Theresa. As I analyzed the original and redo placement session think-aloud protocols, I also focused on the 30 problematic Course 10

students read by Theresa. As I said earlier, Theresa's list of problematic Course 10 students contained 11 more students than Adele since she read 13 students and placed them into Course 11 in the redo session without ever pushing them down, making a different placement decision than their original decision. Only two of these students were participating in the original placement session for there to be a record of Adele's original reading. Using these 30 students in total, I counted how many students Theresa discussed each criterion on and if that discussion was positive, negative, or mixed. Table 10 details the problematic Course 10 students' totals for each of the reading and writing criteria as discussed by Theresa. Even though Theresa is not the expert-reader for the course, her criteria are still important to consider since many of these students had their course placements changed between the two reading sessions.

Table 10

Total Number of Problematic Course 10 Students Discussed by Theresa for Each Placement Criterion

# of Students	% of Students	Name of Criterion
28/30	93.33%	Reading issues/quality
26/30	86.67%	Assignment task
22/30	73.33%	Working with text
18/30	60%	Essay/ paragraphing knowledge
7/30	23.33%	Length/amount of writing
6/30	20%	Language
5/30	16.67%	Grammar
5/30	16.67%	Thinking
4/30	13.33%	Framing/moves
4/30	13.33%	Use of required texts
2/30	6.67%	Academic rhetorical conventions
1/30	3.33%	Print conventions
0/30	0%	Thesis/project
0/30	0%	Writing process

More mixing of positive and negative comments for Theresa. Just like with Adele's comments on these problematic Course 10 students, Theresa's comments overall trended more toward mixed comments that mentioned both positive and negative attributes, with the addition of some positive comments as well. However, this trend was most significant on the two main criteria of reading quality and the assignment task. With these criteria, comments that were once predominantly negative on regular Course 10 students became predominantly mixed. The other criteria showed more slight movement toward a mixing of positive and negative comments, but for many of these criteria—essay/paragraphing knowledge and framing/moves—the majority of comments were still negative. Comments on a student's work with the text were the exception, as they were with Adele, in that the characteristic of these comments remained largely negative. Whereas Theresa commented only negatively on regular Course 10 students' work with the text, on these problematic Course 10 students, she gave a few students mixed assessments.

Interesting differences between Adele and Theresa as readers of problematic Course 10 essays. Since Adele and Theresa were reading with different recent course teaching experience (Smith 1992/2009), it is not surprising that their criteria are different. Figure 4 compares Adele and Theresa's problematic Course 10 criteria, displaying the percentage of students commented on for each criterion.

As I've already indicated, the same four major criteria are the most frequently commented on criteria in terms of overall students—reading, assignment task, essay and paragraphing knowledge, and work with text. However, with problematic Course 10 students, comments on students' work with text outnumber essay and paragraphing

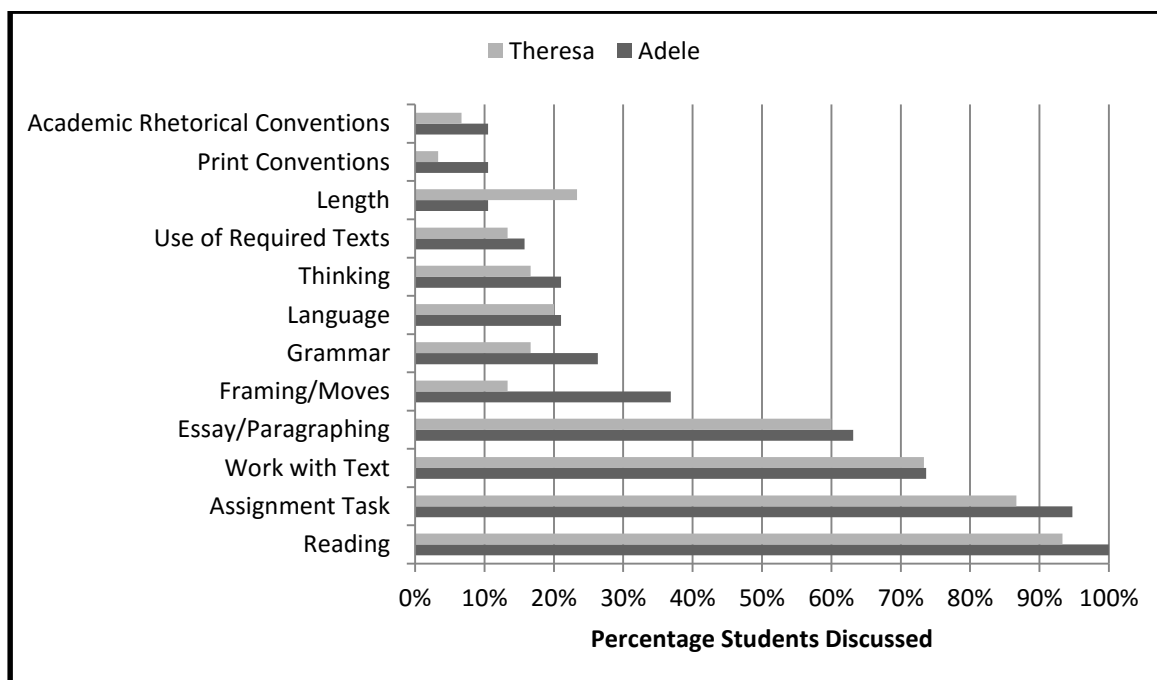


Figure 4. Comparison of readers' problematic Course 10 criteria.

comments for both readers, a ranking not seen for any other student group. As with other student groups, Theresa seemed to comment less frequently overall for the major criteria, but the major criteria of interest for problematic Course 10 students were remarkably similar for both readers. Where the readers differed with problematic Course 10 students was with comments on framing and moves, grammar, and length. Adele discussed problematic Course 10 students' framing and moves and their grammar far more frequently than Theresa, with the difference in comments on framing and moves being striking. Theresa, as she did with problematic Course 11 students, discussed the criteria of length a good deal more than Adele.

Key criteria that present problems for Course 10 placement. Looking at the total number of students commented on for each criterion is a fairly easy way to determine the criteria most important for problematic Course 10. But we once again find

the same four key criteria—reading, assignment task, essay and paragraphing knowledge, and work with text—at the top of both Theresa and Adele’s lists. As I’ve said multiple times before, it is important to look at the characteristics of these comments to have a full understanding and view, investigating the major criteria, those criteria that are different from earlier discussed groups of students, and those criteria that show differences between the readers. Next, I look a little deeper into the qualitative end of the data to see what characteristics of these particular criteria—including their positive, negative, or mixed determination—present difficulty in determining Course 11 placement. I focus first on the patterns in Adele’s reading, and add in a discussion of Theresa’s assessment once I’ve sketched the trends from the expert reader.

Reading issues and quality. The most important criteria for placement overall has consistently been reading. And problematic Course 10 students are no exception. Just as with other student groups in this study—unproblematic Course 11, unproblematic Course 10, and even problematic Course 11—reading was the most frequently discussed comment, and each group of students had particular trends in the kinds of reading comments made. The previously discussed trend in reading with unproblematic Course 10 students was serious or overall misreading. Some of the regular Course 10 students would also avoid reading or using one of the assignment texts, calling into question their ability to read that text at all. As these are serious issues, it makes sense that they would place students into Course 10, the lower course option. While some of the problematic Course 10 students examined here exhibited these characteristics, the number was far less than normally seen in unproblematic Course 10 students, and that makes sense given that these students presented difficulties to the placement readers. For instance, six of the 19

problematic Course 10 students read by Adele had overall misreading, and only three left out one of the texts. Instead, nine of the 19 students showed some misreading mixed with moments of good reading, a characteristic that places a good number of them closer to students on the edge of Course 11 in terms of reading.

While this contrast between student groups is interesting, a far more curious observation about problematic Course 10 students and reading is how often Adele expressed uncertainty about her assessment of these students' reading. For eight of the 19 problematic Course 10 students, Adele expressed some uncertainty as she discussed the student's reading. Having examined these students, it is clear that the uncertainty about these students' reading stems from interference from a number of criteria that impact Adele's ability to judge the students' understanding. Student 1010-01-11 is a representative example.

After first suspecting that student 1010-01-11 in Original Session 6 is "on the wrong track" with the reading, Adele noticed that he had some good reading of how Graff and Pluto connect. She attributed the confusion over how much he understood to multiple other attributes that made it hard for him to express himself and for her to tell how much he understood. First she talked about the limits of his language: "He doesn't have the language to say what he wants to say, but he definitely has some good ideas here. That he's obviously got from the reading." She also credited some confusion to issues with moves where he made "gigantic leaps" in topic, and she mentioned, as she did with many other students, that "there's a lot of unpacking" needed for him to spell things out to the reader by working with the text differently than he's doing. She concluded, "I think there's real potential with this kid, but I think he needs a lot of support. And I think

he would be lost in [Course 11].” The sheer number of other criteria that Adele had to weigh in uncharacteristic ways with this, and many of the other problematic Course 10 students certainly shows why deciding the placement of these students may be more difficult.

Assignment task. Unproblematic Course 10 students frequently switched the assignment task, said too little on the connection they were building in their response to the assignment, or they built their connection between the texts on a misreading. Some of the problematic Course 10 students share these characteristics. Six of the 18 students commented on by Adele for assignment issues were described in some way as not doing the assignment. Half, or three of these six students, had only one text, so they could not do the assignment, and the other half discussed the texts making no point or connection at all. Three students named a connection at the end of their piece, but said very little about it. And two students tried a connection that contained a misreading. While this accounts for 11 of the 18 students in this group commented on for assignment issues, the other seven students tried the task with some success. These students were 1010-01-10, 1010-01-11, 1010-04-11, 1010-04-13, 1010-05-03, 1010-06-14, and 1010-07-03. That means, at least on the issue of meeting the assignment, these seven students looked more like their Course 11 peers than those in Course 10.

Returning to the issue of reading, discussed above, I want to compare these seven students’ assessments of reading to their assessments on the assignment discussed here. Five of the seven students who were described as attempting the task, an attribute more similar to Course 11 students, also had reading characteristics more in line with Course 11 than Course 10, students 1010-01-10, 1010-04-11, 1010-04-13, 1010-05-03, and

1010-07-03 specifically. When we look at the combinations of these criteria, the difficulty placing these students makes sense as, so far, these five students seem to share attributes with Course 11 students on key criteria.

Work with the text. As unproblematic Course 10 students were described earlier, it was found that a few of these students had some trouble with copying, but most had too little work with the texts to show that they understood or too little to detail the connection they saw. While some of the problematic Course 10 students examined here exhibit these same characteristics, there were differences in these students' work with text. Three of the 14 problematic Course 10 students who were commented on regarding how they worked with text had copying issues, but two of these students had unique copying issues. One student copied the moves off the assignment sheet, "basically just [repeating] every part" of the assignment (1010-02-02, Redo Session 3), and one copied bits of language from the texts rather than whole passages, such that Adele was unsure if the student might have understood more from the reading than typical students who copy (1010-06-01, Redo Session 1). Thus, while these problematic students copied, just like their unproblematic counterparts, the copying did not suggest as clearly that they struggle with understanding what they've read as typical Course 10 copying does.

Adele's comments on working with the text in regards to the remaining 11 problematic Course 10 students represent a range of problems these particular students exhibited. All of her comments can be loosely categorized under a sense that the student has done too little with the text, and as such these students were not all that dissimilar from regular Course 10 students, although the lack of work with text for two of the students hurt their ability to come to a full sense of the texts and address the assignment.

One student was described as “not [knowing] how to work the two texts together” (1010-04-08, Original Session 5) and one didn’t do “enough” work with the texts to fully understand and make a connection (1010-06-07, Redo Session 1). These students then represent the bottom of this group. For both students, Adele believed they had some understanding of at least one text—they “have some interesting things,” (1010-06-07, Redo Session 1)—but she thought that they hadn’t learned processes of reading, writing, or thinking that would help them arrive at a strong reading of the two texts together, or “they just don’t know. They just can’t make that move. So they write a nice long essay” and “make connections with real life” instead (1010-04-08, Original Session 5).

The remaining nine students all had at least some work with the text towards a connection. Four of these students started with the text, but then they moved away from that work in favor of their own examples or response, and Adele says these students have to learn to “stick to the text” (1010-01-07, Redo Session 1). Five of these students were described simply as “trying to do a little bit of work with the text,” but with each student Adele talked about how she wanted to see more work with the text (1010-05-03, Redo Session 3). While the details of comments on these problematic Course 10 students’ work with text display similarities to Course 10 students, recall how the criterion of work with the text is related to reading. Adele wants these students to work more with the text largely because their doing so would make her judgment about their reading of these texts clearer, specifically their ability to read one text through another to make a connection. So even though these students are like their regular Course 10 peers, they are like their peers on an aspect of academic writing through which writers present enough evidence to their readers, and as such, what constitutes enough, is at question here.

As a final note, as we have gone through each of the criterion so far for these problematic Course 10 students, I have tracked a set of students with positive assessments made on each criterion so far. Recall that there were five problematic Course 10 students who had both their reading and assignment issues assessed in ways that suggest they were more like Course 11 students. Adding in the examination of these students' work with text issues, four of the students have comments in this category that indicate the student tried at least some work with the texts, even though Adele wanted to see more work directly done. Therefore, four of these students, 1010-04-11, 1010-04-13, 1010-05-03, and 1010-07-03, still have attributes that make them very similar to their Course 11 peers.

Essay and paragraphing knowledge. Unproblematic Course 10 students frequently wrote responses with no paragraph breaks or they had two paragraphs separating the two texts (almost two-thirds of these students who received comments on this issue fit one of these descriptions). The remaining unproblematic Course 10 students commented on for this issue (about a third overall) had multiple paragraph responses, one student who had a full essay shape. Comparing these regular Course 10 students to the more problematic ones discussed here, we find that the full range of these essay and paragraphing issues was still present for problematic Course 10 students; however, the students now broke down differently. Five of the 12 the students commented on for essay and paragraphing by Adele had no paragraph breaks or a two paragraph response that separated the texts, and the other seven had multi-paragraph responses, two described as having an essay shape.

Returning to the problematic Course 10 students who show signs of being the most problematic for placement decisions, those students I've tracked as having a more

positive assessment on each criterion discussed so far, all four of the students with positive assessments on reading, assignment task, and work with texts also had strong comments here on essay and paragraphing—that is, they each were described as having a multi-paragraph response. These students, 1010-04-11, 1010-04-13, 1010-05-03, and 1010-07-03, have attributes that make them very similar to their Course 11 peers now on all of the four major placement criteria. I will discuss these students more in-depth shortly, when I factor in Theresa’s comments on each.

Framing and moves. Unproblematic Course 10 students were rarely commented on by Adele for their issues with framing and moves. When they were though, the comments suggested the students seemed not to have planned or were primarily planning unhelpful moves of summary. The problematic Course 10 students, in contrast, were commented on more frequently for this criterion than their regular Course 10 peers, but what was said about them was very similar. Four of the seven students commented on for issues with framing and moves had comments suggesting a lack of planning or an overuse of summary, such that it drove the structure of their response. One student of the seven was praised for the inclusion of an introduction. The final two students commented on for this criterion had an issue with quick and unexpected changes in topic, what Adele characterizes as “leaps” (1010-01-11, Original Session 6), and as such they represent a new issue we haven’t seen previously—i.e., the student struggles to plan and signal moves for the reader.

Grammar. Unproblematic Course 10 students showed serious errors in grammar of a nature that either interfered with their meaning or were significant enough, particularly in terms of ESL language learning, that the reader thought the student would

benefit from the additional instructional time offered in Course 10. When we look at the more problematic Course 10 students, students who were not easily placed in Course 10, these students were primarily commented on for the same issues—serious errors that interfered with meaning. Perhaps more striking though was that only slightly more than 26% of these students were commented on for grammar issues. While almost 54% of regular Course 10 students were commented on for grammar issues. In short, these students display fewer grammar issues than their regular Course 10 counterparts.

Adding in Theresa's comments on the most problematic Course 10 students. As I have discussed each of the major placement criteria above, I have focused on Adele's criteria and think-aloud record, as she is the expert Course 10 reader. However, now I want to look at Theresa's comments on problematic Course 10 students. Given that these students must be pushed down by the Course 11 reader to be read by Adele, Theresa's think-aloud discussions of these students are available for comparison. I begin by looking more closely at what Theresa said about four students specifically. As I have discussed Adele's problematic Course 10 criteria, I have continually returned to and built a subset of four problematic Course 10 students read by both readers who appear to be more problematic than others. I have shown that these students have characteristics on more than one of the major placement criteria that make them appear more like their Course 11 peers, at least in regards to Adele, but now I want to look at what issues most concerned Theresa as she pushed them down.

Interestingly, these four students were all tweeners, pushed down by Theresa, back up by Adele, and eventually discussed in a conversation between the readers where they determined placement together. Two of these tweeners were put into Course 10, just

as they were on their original reading, but two were placed into Course 11, a different decision than was made on their essays originally. I want to look more closely at this subset of students then to consider why they were not placed into Course 11 by Theresa if they looked more like Course 11 students to Adele, hoping to shed light on what is at issue with many of the students in the problematic Course 10 group. All four students, 1010-04-11, 1010-04-13, 1010-05-03, and 1010-07-03, were described as having a mix of misreading and good reading, but they were attempting the assignment, and at least writing a multi-paragraph response. However, these students, it was suggested, needed to work more with the text.

Examining Theresa's think aloud reading of some of these students reveals that for many, she was torn as well. On 1010-07-03, in particular, she expressed doubt multiple times as she was reading, saying, "I don't know about this essay" (Redo Session 2). Looking just at the criteria she discussed with this student, Theresa noted how the student didn't mark moves so that the paragraphs—there were multiple—seemed like "separate entities" (Redo Session 2). She mentioned how the student juxtaposed two ideas from Graff, but didn't explain as she should. And Theresa noted that the student had trouble fully achieving the assignment and making a connection between the readings because she didn't really make moves and explain connections throughout. Theresa did not discuss the quality of the student's reading as a separate consideration. On criteria alone, this student seems more like a Course 11 student.

But Theresa was still concerned about student 1010-07-03. The student incorporated her own example about the election of Obama, and Theresa was unsure how that connected to the readings or assignment, as the student wasn't clearly explaining the

connections. As she thought about it, Theresa said she could “see that there can be a connection between those two things, but there isn’t one that’s made” (1010-07-03, Redo Session 2). Shortly after these words, Theresa claimed it was a “Course 11A” paper, and then she said she couldn’t put the student in Course 11. I conclude from this that the biggest impact on Theresa’s decision here was her concern with how much she’s reading into what the student said as she can hypothesize a connection that the student hasn’t done the work to make herself.

Theresa pushed the student down, and Adele went on to read the same essay and express concern that the student “lose[s] it” when the student focused her reading of Graff on sports and how sports have “rule books” (1010-07-03, Redo Session 2). The idea of rules ended up in the remaining paragraphs discussing Pluto and the Obama election. Adele said, “I mean, if they didn’t have this rule book issue, this would be a not bad essay,” but she also said, “I mean like I don’t know.” She ultimately decided to push it back up as the student “just like [gets] off base there on one point,” but had so much else going for her (Redo Session 2). As they met to decide placement together, they both expressed their separate concerns that the essay was hard to follow and that the rule book thing was “a moment of obsession” (Theresa, 1010-07-03, Redo Session 2). They read through the whole piece together and characterized the student as reading in Course 10 ways, where Course 10 students like her get “attached to a reading,” but also not being “so far off” the Course 11 students on the other writing based structures like “paragraphing” and “sentences” (Theresa, 1010-07-03). Together, they decided Course 11 placement.

This one student reveals several important points that characterize both problematic Course 10 students and issues of expert reading for Cardinal College. What we learn here in terms of the problematic Course 10 students is two-fold. First, these students truly have mixed abilities—some abilities that look like Course 10 and some that look like Course 11. They are indeed difficult to place because, with only one piece of writing to look at, it is hard to decide if they will struggle or not. Second, reading, as a criterion alone or in its relationship to the other criteria, is what is primarily used to differentiate between Course 10 and Course 11 students. Examining the other three students in this subset bear out these two observations. Despite looking like a Course 11 student in ways already described, student 1010-04-11 doesn't "use the reading" of Graff to look at Pluto as strongly as Course 11 students do so she was again placed in Course 10 (Adele, 1010-04-11, Redo Session 1).

Student 1010-04-13 was as strong a writer as Course 11 students, but only got her reading to work for her nearer to the end of her piece since she over focused on her own ideas on education rather than using the text for most of her essay. The kind of revision that would be required of this student, where she must throw away most of her draft and "start all over" with her reading is "harder" to adequately deal with in Course 11, so the student was placed into Course 10 (Theresa, 1010-04-13, Redo Session 3). The discussion on student 1010-05-03 reveals the biggest issue to be that the student didn't do enough work with the text, which made it hard to determine whether her reading was a cliché (as it would be if she had approximated the text as a Course 10 student) or whether she had an understanding that she was just not showing through close textual work (as a Course 11 student would need) (Redo Session 3).

Yet in addition to the truly mixed nature of these students' abilities and the primacy of reading as a placement criterion, we also learn something about course teaching expertise as it relates to Cardinal College by looking at student 1010-07-03. The previous existence of a course between the two current courses, Course 11A, which has since been discontinued, still impacts placement readers' thinking. Theresa first decided student 1010-07-03 belongs in Course 11A. Then she made a decision about what to do after that. As I've already discussed, many of the students I flagged on the problematic Course 10 and problematic Course 11 lists are on these lists solely because one or both readers said that they were old Course 11A students or said they were described as "borderline" (1010-05-03, Redo Session 3). While these designations happened more during the redo session (31.48% students of the 108 participating were flagged as Course 11A) than they did during the original sessions (20.59% of the 108 participating students were flagged as Course 11A in their original placement reading), it is clear that even in the original reading, the understanding of what a "Course 11A" student is still stands long after the course has been discontinued. As I go forward, addressing other subquestions of the research question, I will talk more about why this old course still has such an impact on placement, and what that means for Cardinal College's expert-reader system and the adequacy of its placement decisions, particularly as it pertains to the most vulnerable set of students—the problematic students described here.

Considering those students with different decisions only read by Theresa in the redo sessions. As I said at the start of my discussion of problematic Course 10 students, many of these students were not participating during the original placement sessions when their essays were officially placed into Course 10, but their essays were all included

in the redo sessions and were reread. In these redo sessions, 15 problematic Course 10 students were now placed into Course 11, and 13 of these students were placed in Course 11 during the redo session without ever being pushed down to be read by Adele. An additional two students were never pushed down in the redo session, but original session data exists when they were placed into Course 10 and were read by both readers. This allows for a comparison of Theresa's first and second readings, as well as a comparison of these readings to Adele's original reading.

Student 1010-01-11, when originally placed, was described by Theresa as writing a response that was "so deeply summarized and not interpreted," that while she thought he understood the text, she couldn't decide if he simply needed to be told to unpack or would require "so much practice" in the "intellectual moves" of "spell[ing] it" out for the reader (Original Session 6). Being cautious, she pushed him down to be read by Adele even though she thought he may be pushed back up. For Adele, there was also the suspicion of misreading, but she concluded "he doesn't have the language to say what he wants to say" (Original Session 6). Since he also was described as only just attempting the assignment connection, having spelling and grammar issues, using vague language, and making "gigantic leaps" in topic, Adele concluded he "needs a lot of support" and "would be lost in Course 11" (Original Session 6).

During the redo session, 1010-01-11 was again read by Theresa. This time she decided to place the student into Course 11 because she believed he understood the "meta-argument" of the text without the details, and needed close reading and essay lessons (Redo Session 1). What the example of this student shows is the primacy of reading for Cardinal College placement decisions, yet again. The difficulty placing this

student is captured by comparing Theresa's first and second readings. Does this student have a reading problem that will require practice and instruction on how to think it through, or does he just need to read more closely and unpack the text to make his strong general understanding more detailed? For placement readers, this is a key distinction.

Course 11 students, as I discussed, are expected to have some misreading issues, but they are reading issues that will solve themselves when the student learns to use the text to show a reader what they see. The implication is that they will, as they choose and explain quotations, come face to face with their own misreading and solve it. But Course 10 students will struggle even if they are sent back to the text to document their reading. They will not be sure how to do this close-reading work, or the habits that led to their misreading will persist, and they will need continued practice and a good deal of guidance. The key difference, it seems, is that a Course 10 student will need help and support—the additional credits and time—while a Course 11 student will work out their misreading.

Student 1010-07-06 also had a different decision made about her placement essay, and we also have both original and redo session data on this student. During her initial reading, Theresa assessed this student's response as being off assignment since it only discussed Pluto, a facet that she believed was a "reading thing" since the student likely avoided the more difficult text on purpose (Original Session 5). Further, she said that "this is a kid...who's been rewarded for turning in assignments even if they don't match the prompt, thinking...that it doesn't matter" (Original Session 5). She pushed the student down, and Adele read it. Adele had trouble at first seeing what the student wanted to do with their chosen information, and she said the student had lots of summary, summary

that picked out important ideas from the text, but also did nothing with the information. It was at this point that Adele discovered that Graff was missing, and she said the student had an essay, but was “following some sort of format...where you don’t have to work with two texts” (Original Session 5).

When rereading this essay during the redo sessions, Theresa now described this student as using Graff, but indirectly, as she noticed a line where the student referenced the “true meaning of having an education” (Redo Session 3). But, as she described it, this “little eentsy, teensy, tiny bit of Graff” was embedded in a summary of Pluto that was “presented as an essay” with little error (Redo Session 3). Having said all of this, she reread the piece to herself briefly and in an uncharacteristic move, directly addressed me. She said:

Oh, yeah, yeah. I’m gonna say I can put this in. And I think it would be in. If I say not in, then Adele can say it’s in, and we can note that it’s problematic. Do you have any preferences to which I do? (Theresa, Redo Session 3)

I reminded her that she had to make a decision as if I was not present, and she concluded “Yeah. I’m gonna put it in,” saying it was “an example of a Course 11A.” This student’s record gives more interesting things to consider, both about reading and about the interference of the old Course 11A class. Between Theresa’s two readings, the biggest change has to do with how she assesses the student’s relationship to the required texts.

In the first read, Theresa didn’t see Graff, and believed that the student couldn’t read it, so avoided it. But, in the second, she noticed one phrase that suggested the student had taken in Graff, at least so far as to know that education was under consideration, and the issue became more complex. It was no longer an issue of just reading, as this one

phrase suggested that the student's thinking did take in some of the ideas of the text. At that moment, not much else was said about why Theresa decided as she had, but the last point Theresa mentioned was that this was a Course 11A student. When it existed, Theresa frequently taught Course 11A, far more frequently than she taught Course 11, and she still describes herself in those terms and still makes some placement decisions influenced by what she knows about teaching those students. The problem with the influence of Course 11A is obviously that it no longer exists in an official way. As I will discuss when I turn to the second subquestion of research question 1, officially there is no Course 11A, but since there are teachers with prior Course 11A experience, and those with none or little, there is a kind of unofficial Course 11A. However, since students aren't guaranteed to be placed with these teachers, Theresa's decisions based on these attributes is potentially problematic.

There were 11 other students who Theresa no longer pushed down in the redo session, but who she placed into Course 11, making a different decision than was originally made on these students. None of these students were participating in the original session so as to have think-aloud data, but reviewing the designations made in notes from the original session, Theresa called three of these 11 students Course 11A students (1010-05-09, 1010-06-08, and 1010-07-05). And this designation means her reading was affected by her knowledge of that old course. This is not necessarily a bad influence, as the data we have considered so far does not yet show which of the two placement options was the better decision for these students, but it does show that Theresa's expertise as a Course 11A teacher is at work in helping her differentiate between borderline students. Checking the list of designations again shows a fourth

student, 1010-02-03, who was a tweener in the original session, so it isn't shocking to find a switch with this student's decision.

On the remaining seven students, Theresa directly discussed the problem as a lack of close reading on four essays (1010-01-08, 1010-04-10, 1010-06-09, and 1010-07-08). As I discussed earlier, students' essays with close reading issues are difficult to place because the lack of close reading in an essay can hide a student's larger reading problems or can simply be the student not knowing how to show his or her reading. Determining which is which is sometimes difficult for the placement readers. On two other students, what Theresa said expressed her own uncertainty. On 1010-05-06, she verbally debated about the issue of the student having written two separate essays, with the readings unconnected, as possibly being both a Course 11 and Course 10 problem (Redo Session 3). And with student 1010-06-03, Theresa hesitated when deciding: "I'm saying. I'm saying uh. I'm saying this is in though" (Redo Session 3). The hesitation suggested Theresa was not certain, even though she didn't verbalize that uncertainty or an explanation on exactly why she made the decision she did. On the final student, 1010-07-07, I believe Theresa's decision was affected by the conditions of the room. First, as she began reading the student's essay, she says "I may have to go down and get a cool drink in a minute," in reference to how hot the room was. Next, the amount of thinking-aloud given on this student was far less than the other students in the session. And finally, immediately after she finished the decision for this student, she paused the tape to get a drink. (Redo Session 2).

Taken together, these students with changed decisions show that problematic placement students have characteristics that make them difficult to determine—they have

attributes that look like Course 10 and they have ones that look like Course 11. But also, this subset of students shows that differentiating between problematic students comes down to the issues of the student's reading and how the student works with text to come to a reading of the two texts together.

Findings on the Expert Readers' Placement Criteria

The primary purpose of detailing so extensively the criteria on which placement decisions are made for the various courses is to have a comparison point for course teachers' criteria. This study purports to directly compare an analysis of the placement criteria to those of the course, but to do so we must first understand what criteria are involved for each course. While I will sketch the differences between Course 11 and Course 10 students on each of the major placement criteria shortly—what their reading looks like, how they typically address the assignment, the ways they work with text, and their knowledge of essays and paragraphing—before I detail that sketch, I want to reiterate two important findings from this analysis of the placement think-aloud data.

First, the differences and overlap in the way each criterion was discussed with Course 11 placement students, Course 10 placement students, and both sets of problematic students suggests that each criterion exists on a continuum of competency, and my sketch of each criterion therefore will show the differences this way. Second, the particular attributes of each separate criterion of placement were seldom considered separately, even though that is how I will sketch them. Rather than being considered separately, these criteria were weighed in relation to what they suggested about other criteria, most importantly what they suggested about a student's reading.

I believe the placement readers typically weighed criteria against each other because their focus was on making a contextualized teaching decision about a student's ability to handle or not handle the college-level course (Huot1996/2009). They were not interested in a grading, ranking, or rating decision. Thus, they were not reducing their view of the student to a measure—they were attempting to contextualize their view through the consideration of evidence that might suggest a student would be able to make growth as an academic reader and writer in a fifteen week, three-credit class that provides particular instruction, or instead would require a six-credit course that provides more time and scaffolding. This strategy—a strategy of considering criteria together over, for instance, rating essays or excluding students from the mainstream course who lack specific criteria outright—was both what helped readers use their expertise to determine where students existed on a continuum of acquisition and what made determining placement infinitely more complex when presented with mixed evidence.

The most important criterion for placement was reading. An analysis of the students placed in each course revealed that placement readers noted three aspects of reading quality: overall misreading of a text or texts, important ideas from a text understood by the student, or specific moments of misreading mixed with moments of understanding. Seeing how placement readers connected their observations on reading with other criteria, for example how problematic students' work with text affected their reading of the text, we see the difficulty in placing students with mixed reading skills centers on not being able to see enough to know what will happen when their course requires them to pay close attention to and use the text directly. Thus Figure 5 presents the reading continuum with two different ways of seeing students with mixed reading in

the middle and a placement decision often being made between those two points on the continuum.

Reading			
Misreads completely, offering what readers call fake reading, making things up, approximation, or clichés	Has some misreading and some understanding of important ideas; the student is thought to need support and practice at close reading	Has some misreading and some understanding of important ideas; the student will be able to unpack fairly easily and fix misreadings	Largely understands readings

Figure 5. Reading issues continuum.

The next most frequently commented on criterion for placement readers was how well a student understood and addressed the assigned task of the placement prompt. As readers thought-aloud on students' responses to this prompt, their comments were categorized into one of three main types: students who did not address the assignment at all, students who successfully addressed the assigned task, and students who at least attempted the task, with varying degrees of success. As with reading, the varying degrees of success with the assignment had to be analyzed by the placement reader to determine what was suggested about the student's ability or inability to read at the level of analysis, and how it was related to their direct or indirect use of text. Figure 6 presents the assignment continuum with two different ways of seeing students who attempted the assigned task in the middle and a placement decision often being made between those two points on the continuum.

On the issue of essay and paragraphing knowledge, the third most frequently made comment by placement readers, readers were concerned with how the students' responses showed their understanding of classic essay structure and, failing that, how the

Assignment			
Changes the assignment; uses only one text	Tries assignment but connection very unclear or built on some misreading; may use one text indirectly	Tries connection with some success; may use one text indirectly	Understands & does the assignment; uses both texts

Figure 6. Assignment continuum.

students' responses demonstrated paragraphing knowledge like breaking to new paragraphs for new ideas or points. There were four descriptions of students on this issue: some students had one large paragraph (or perhaps two separate paragraphs that never mix the texts), some students had multiple paragraphs, some students had multiple paragraphs with an inclusion of some classic essay moves, and finally, some students had identifiable essays that also included a clear sense that the student was making purposeful and preplanned moves. What is important to point out here is that rather than make placement decisions at the midpoint on this continuum, many students who just had multiple paragraphs would end up in Course 11, and on occasion a few with even only one or two paragraphs separating the texts would as well. Figure 7 presents the essay and paragraphing continuum.

Essay and Paragraphing			
1 big paragraph or 2 paragraphs/responses separating the two texts	Has multiple paragraphs	Has multiple paragraphs and some sense of essay moves; but may lack full essay	Essay structure and sequence of moves

Figure 7. Essay and paragraphing continuum.

The fourth most frequently commented on criteria by placement readers had to do with how a student worked with the text or texts. The major concern of readers with this issue was how students used the text—choosing to summarize or quote—but since quotation was preferred, readers were also concerned with how and how much students worked with quotations— marking quotation boundaries, explaining them, and connecting them to their points. On the explanation of quotations, readers frequently discussed how students needed to work more with quotations to explain them, signaling either that the reader felt the student didn’t work enough with the text directly to fully come to a clear sense of what they were saying themselves or they hadn’t worked with the text enough to show the reader what they were saying. Figure 8 presents the continuum on working with text, and the important distinction for placement is represented by the midpoint as readers attempt to determine if the student has a reading they are not showing, or if they will need time and practice to develop readings of complex texts.

Work with Text			
Copies significantly from the text without marking and primarily summarizes ideas from the text	Summarizes, answering the assignment either without quote(s) or without explanation of quote(s); student must work with the text more to really come to a reading	Summarizes some but also has quote(s) lacking explanation that could be useful for stated points; student must work with the text more to show their reading	Has useful quote(s) and tries some explanation or close reading of them, connecting them to stated points

Figure 8. Work with text continuum.

Having now sketched the four major criteria, and the specific attributes that distinguish Course 10 placement from Course 11, I want to sum up a major finding from my analysis of the placement think-aloud data. There is an uncertainty about what the exact border is of each Course 10 and Course 11 and how much “space” exists between them because of two important factors. First, the boundary of the courses is primarily based on the potential for a student to experience reading difficulty, and that potential is hard to determine when the single piece of student writing has characteristics that present readers with insufficient evidence about the student’s reading. The limits of one piece of writing are well known in the literature, which is why the use of portfolios is suggested (Conference on College Composition and Communication, 2009). However, Smith (1992/2009, 1993) believes, as do others (Haswell & Wyche-Smith, 1994; White, 1995a, 1995b) that it is practical and possible to make adequate placement decisions on the basis of one piece.

For Smith (1992/2009, 1993), the possibility of adequate decisions was related to teacher’s most recent course teaching experience and how well they used their course teaching knowledge. And this relates to the second factor causing uncertainty about the borders of Cardinal College’s two courses. Placement determinations are influenced by an old course, Course 11A, that was discontinued by the administration in 2008 and previously served the needs of students who had fairly competent writing, but questionable or problematic reading. It isn’t clear yet, from the data examined thus far, if expertise on this old course, admittedly no longer recent course teaching experience, helps readers decide placement, or if it interferes with placement decisions. Since the curriculum has changed, holding on to this Course 11A distinction during placement

could possibly help readers decide what is best for these old Course 11A students on a case-by-case basis inside the new curriculum, or it could interfere in their placement decisions and potentially prevent teachers in the writing program from developing a new sense of the courses, a new expertise, that accounts for Course 11A students as part of their population—that is, using Course 11A knowledge at placement may be helping the program match problematic students to the best course for them, or it could be keeping the remaining courses in the program from adapting adequately to a shift in population.

Going forward, I consider Smith's (1992/2009) claim that expert placement readers use privileged course knowledge by examining the placement think-aloud protocols for this knowledge, and I investigate how old Course 11A distinctions influence readers' privileged course knowledge. Then, I turn to look at the course phase of the data, as teacher's recheck the placement decision during the first week of the course.

Are Expert-Readers' Placement Criteria Teaching Specific?

Focusing again on just the placement sessions—both the original and redo—I examined the transcripts for instances of what Smith (1992/2009) called privileged course knowledge to consider if placement criteria are teaching specific. Smith posited that expert-readers draw on their experiences teaching students of the course to reflect on the students behind the essays and consider the teaching of real students when making their placement decisions. As I coded and analyzed the placement think-aloud transcripts, I looked specifically for examples of privileged course knowledge. In what follows, I present the two types of privileged course knowledge I uncovered in the transcripts. Additionally, I also discuss the discovery of two other kinds of comments made by readers that seem related to the readers use of privileged course knowledge, but which

introduce potential interference to the expert-reader placement model. After discussing the use of privileged course knowledge and its benefits, I consider the meaning of these other comments.

Privileged Course Knowledge

My analysis of the placement transcripts revealed that expert placement readers at Cardinal College do indeed use privileged course knowledge. Of the 26 participating students in the original placement sessions, six of the students, or 23.08% had one or both of two kinds of privileged course knowledge comments. In the redo sessions, when there were 108 participating students, 30 of the students, or 27.78% had privileged course knowledge comments of one or both kinds. Furthermore, the instances uncovered fall into two distinct categories of comments: understanding where students may struggle or excel in the course and knowing how to teach course students with specific issues. In what follows, I present the discussion of these two categories of privileged course knowledge and my resulting conclusion that expert placement readers indeed use teaching specific course knowledge to place students because such knowledge is particularly helpful when making a placement decision on the most problematic students.

Understanding where students may struggle or excel in the course. Placement readers often see the student behind the essay, recognizing the student as a student and more than just a set of criteria. Seeing the student behind the essay means readers can note where particular students may struggle or excel in their course and how they might behave in reaction to their struggle or ease, and knowing this takes privileged course knowledge of the kind Smith describes. To know where a student will struggle or excel, teachers must have experiences teaching real people in their course.

In the original placement transcripts on the participating 26 students, only two students have such comments, which at 7.69% of the students overall does not seem significant. However, during the redo sessions, when 108 students were participating, 21 of 108 students, or 19.44%, had such comments on their essays. This larger portion during the redo sessions suggests that this kind of course knowledge is more frequent than the original session data, with its low number of participants, can show. I offer a few examples of comments that show expert-readers understand where real students will struggle or excel in their courses to illustrate how seeing the student behind the essay and imagining them in the course helps readers determine course placement.

In Theresa's initial reading of Student 1011-03-06, a tweener in Original Session 5, she noted that the response was one "lumped up" paragraph with unexplained quotations, making the piece read more like a "fill-in-the-blank answer." Adele, also saw the lack of structure, but believed that there was "good thinking going on here" that was "based on what they've read" (Original Session 5). Having tweened the student, they reconvened to make the final placement decision together. As part of that discussion, they turned to how the student would react to teaching with Adele saying, "My guess is that she's probably somebody who might. Who would respond to instructions. Like you get these kids who want to learn. They want to do well." Theresa agreed that "she's gonna need someone who's going to direct her to do things." Having made these observations on the kind of student they imagined she would be—an active and willing participant in her learning, but requiring direction—they returned for a second look at her essay and decided that "you know she needs to unpack. She needs to understand how to make her reading visible" and that she would get and respond to that kind of direction in Course

11. Adele and Theresa's visualization of this student as a real person with real reactions to teaching helped them interpret what they saw in her writing and where best to place her.

A similar situation happened with student 1011-05-03, an unproblematic Course 11 student. Despite being unproblematic, in terms of this study, Theresa's think-aloud record from Redo Session 1 shows that she made this student's placement decision by likening him to actual students in her course. While reading his essay, she said the student "[got] the meta-idea of Graff" and saw "Pluto as an example," but he had "not unpacked," made "no references to the text," and had "an example that [didn't] work" for Theresa. So again, we see a student who has issues with unpacking, one of the key criteria that makes differentiating between students difficult. As Theresa tried to decide what to make of this student's issues, she said "I have a student right now that profiles like this, and I think I've learned something." As she continued, she mentioned the "hard part with that student is working with him to understand he has to unpack." In essence, by likening student 1011-05-03 to a student in her class with similar issues, Theresa was able to decide that this student belonged in Course 11. Understanding this student's problem in these terms, Theresa was able to conclude that this student's reading and writing issues could be addressed by Course 11.

Knowing how to teach course students with specific issues. In the natural course of reading placement essays, readers frequently mention how they would teach or handle students with specific reading and writing problems, mentioning specific teaching strategies like the comments they would make on a student's paper, specific revision advice they would give, specific lessons in the course, and even how students would react

to such advice. In the original placement transcripts with 26 participating students, five students, or 19.23%, had moments where placement readers discussed how they would teach a student. However, during the redo sessions, when 108 students were participating, 11 of 108 students, or 10.19%, had such comments on their essays. As I did in the section above, I offer a few examples of comments here that detail what it looks like when readers discuss how they would teach students, and how that knowledge helps reader's use teaching-based observations to make their placement decisions.

Reading the response by student 1011-03-07 in Redo Session 2, Theresa described the student's essay as having "unhooked from Graff entirely," where the student was "gesturing towards Graff but he's not there." Further reading exposed that the student had gone on to see an odd implication in the connection they'd made between Graff and the Pluto demotion, and Theresa linked this to the reading. Once she finished the entire response, she decided that the student's "reading is invisible at this point," but that "a [Course 11] teacher can deal with that fairly quickly actually." While she doesn't name specifically how a Course 11 teacher will "deal" with invisible reading, here Theresa was using her privileged course knowledge to think about how teachers of her course handle such instances like this student's essay. Drawing on this knowledge, she was able to determine if this student's problems presented a normal or abnormal difficulty to the teacher of a Course 11 class.

We see Theresa used similar knowledge regarding how to teach certain issues as she discussed student 1011-03-02 in Redo Session 3. With this student, she noted that the writing was really "two separate essays. One about Pluto" and "one very short one about Graff." In response to the student's separating their discussion of the two texts from each

other, she explained that “in years past” this student would have been placed in Course 11A because “this is a person who doesn’t understand how to bring these two operations together,” who doesn’t understand the “critical operation.” And she said, “Now I have to look at it, and I have to decide is this person going to be able to do Course 11” (Redo Session 3). As she discussed what she saw, she decided that she “would have to say yes” they could do Course 11 because

I could say this in like a comment on ten lines, probably, that you need to bring these two things together. And how do you do that? And then show them how to do that. And then that person might get it. This person is gonna struggle, but I’m gonna put them in rather than down. (Redo Session 2)

As she thought about the work of teaching reading and writing in her course, specifically what she typically says to students and how she comments on their work, she decided that she would place this student in the course. Reflection on the work of teaching this student was what helped her decide, and presumably writing a comment like the one she described is a common experience in her teaching of Course 11.

The relationship of privileged course knowledge to problematic students. As a final observation on privileged course knowledge, before going on to consider other kinds of comments I’ve discovered, the data shows that there is a relationship between privileged course knowledge comments and the placement of problematic Course 11 and Course 10 students. Placement readers at Cardinal College more frequently discussed privileged course knowledge on the essays of students who were problematic placement students—tweeners, Course 11A students, and students who had two different placement decisions made about their essay across two different readings. Of the six instances of

readers using course privileged knowledge during the original placement sessions, five were on problematic students, or 83.33%. Of the 30 instances from the redo placement sessions, 18 were on problematic students, or 60%.

What this suggests is that with difficult students—those students who do not fit neatly into the curriculum—readers will more frequently use teaching-specific knowledge to weigh attributes of students' essays against their knowledge of the typical students in the course to determine placement based on teaching and learning specific criteria. Yet, since some of these students had two different placement decisions made about their essay across two readings, 18 out of the 108 students, or 16.67%, it seems that even with the use of course privileged knowledge, sometimes a different decision is made about a student's placement. Why? Smith (1992/2009, 1993) theorized that some students will never quite fit neatly into our courses, and that is likely a factor here. But additionally, Cardinal College has to consider that old Course 11A designations may be interfering in placement decisions rather than permitting Cardinal College to truly use the expert-reader model. And by extension, as I will discuss later, this interference may keep the curriculum from changing, to account for the loss of the Course 11A course.

The Problem of Course 11A Interference

As I was reading the placement transcripts specifically for privileged course knowledge, I took note of two other kinds of comments made by placement readers that seemed to involve some knowledge of the course. As I analyzed these pieces, I realized that they were strongly related to the previously discussed kinds of privileged course knowledge and to each other. The first type of comment recognizes and identifies students who do not fit neatly inside the boundaries of the courses, what I have already

described in my earlier discussion of the placement criteria as references to the old Course 11A placement option. Sometimes the readers described students as “borderline” or “not fully prepared” for their course, but mostly readers used the label of Course 11A, particularly when they felt a student would have been placed, in previous years, into this discontinued course that fell between Course 10 and Course 11.

The second type of comment I discovered had to do with a reader making a recommendation on a student’s essay for placement with a particular teacher or teachers. These comments were almost always made by Theresa, the Course 11 reader, who is also the director of the writing program. I go on to show there is a relationship between these teacher recommendations and students determined to be on the boundary or old Course 11A students, as Theresa often chooses a teacher based on her sense of the student as Course 11A student or not, and based on her desire to assign the student to a Course 11 teacher who used to teach Course 11A or who did not. Using both these types of comments, I discuss the potential for course boundary knowledge to help readers differentiate between problematic students and make adequate placement decisions, but I also consider the potential for the existing course boundary knowledge to interfere with adequate placement decisions as it focuses on a course that no longer exists.

Recognizing students on the boundary of the course. As I have already discussed when analyzing the placement transcripts for course criteria, placement readers often mention fit-to-course problems with certain students. For the most part, readers use the label of Course 11A when they feel a student would have been placed in the discontinued course that used to exist between Course 10 and Course 11. Course 11A, which no longer exists, was four credits and counted identically in requirements as

Course 11. Despite counting identically, it fell between Course 10 and Course 11 as it was expected to help students who had additional issues with reading and grammar, but not of the severity seen in Course 10 students. As a result, it only required one additional credit than Course 11. The suggestion, when a teacher mentions this designation now, is that the student displays attributes that are some fit to the teacher's course, but that the student does not fit neatly in some way as well.

Understanding which students fit into the course and which do not takes course knowledge, as the reader must recognize the boundaries of their course and student attributes that typically make it up. In the original placement transcripts with 26 participating students, nine students, or 34.62%, had moments where placement readers considered if they thought a student was on the boundary of their course. And, during the redo sessions, when 108 students were participating, 36 of 108 students, or 33.33%, had such comments on their essays. The majority of these comments were made by Theresa, although occasionally Adele would comment in this way as well. Theresa commented on eight of the nine original session students, and one student was commented on by both Theresa and Adele. During the redo sessions, 22 of the 36 students were commented on by Theresa, six by Adele, and eight by both readers. In what follows, I illustrate some examples of these comments, arguing that they are a type of course knowledge, but I also consider the possible problems that are introduced into the placement process when making reference to a course that no longer exists.

Before offering new examples, I want to return to a few students already discussed in the subsections immediately preceding this one. First, student 1011-03-02 was offered above as an example of Theresa's knowledge of how to teach certain course

issues. The student wrote two separate pieces and needed to learn to bring them together. In addition to Theresa's knowledge of how to teach this student, with 1011-03-02 we also see that Theresa recognizes the attributes of students who fall on the boundary of Course 11, as she said the student would have been placed into the old Course 11A because it was common for such students to struggle to do the more "critical operation" of reading one text through another (Redo Session 3). Student 1011-03-02 also shows Theresa's concern with making sure these students are carefully placed, as it is precisely because this student doesn't fit that Theresa turned to her teaching knowledge to make the decision.

But additionally, this student also reveals that Theresa often makes a placement decision on Course 11A before she makes a Course 11 decision. For instance, when Theresa said, "Now I have to look at it, and I have to decide is [1011-03-02] going to be able to do Course 11" (Redo Session 3). This was not an isolated incident either. Theresa frequently thought-aloud about a student's status as Course 11A, often explaining why she would label the student as belonging in that course before she made her placement decision, although the label of "Course 11A" never happened in the original placement sessions. Instead Theresa used phrases like "not a clean 11" (1011-03-02) or she described the student as being under Course 11 but over Course 10. Of the 29 students flagged by Theresa for not fitting the course properly during the redo sessions, she named the overwhelming majority, or 20 students, as Course 11A. On about 13 of the 29 students, her determination of Course 11A preceded her ultimate decision. And, on at least seven of the 29, she directly mentioned that she had to figure out what to do when

she couldn't place the student into a Course 11A, or she agonized over not having that course as an option anymore, since it would best meet the student's need.

It happens as well with student 1011-05-04 when Theresa said first that the student was "underprepared for [Course 11], but not so grossly underprepared as to need six credits" before deciding to place the student into Course 11 (Original Session7). Student 1011-05-04 also reveals part of what is at stake for Theresa in these moments. Since the program has lost the four-credit option that used to exist between Course 10's six-credit load and Course 11's standard three-credit one, Theresa was often frustrated by feeling a student needed more than Course 11, but not as much as Course 10, and she struggled to justify the choice of a six-credit option when her experience meeting these students' needs previously with a less credit-heavy course had shown her the six-credit option wasn't always necessary. In a way, Smith (1992/2009) supports her as he claimed, when discussing teachers' reluctance to move students they feel are low, that his teachers were not wrong about their ability to catch up students who were somewhat below the course.

Theresa's frustration with not having a Course 11A option is understandable. The decision to do away with the course was not made for curricular reasons or because of a change in student population. Instead, the College was going to be experiencing space concerns for a number of years due to the sale of some property and loss of a building, and the administration asked that departments eliminate four-credit classes so all course offerings could occupy the same time slots and be scheduled more easily. But even though the program would lose the Course 11A course, that didn't mean that it wanted to

lose the understanding of the student population that was gained over years of running and teaching that course, hence the continued use of the Course 11A label.

However, Smith's (1992/2009, 1993) expert reader model says that readers must have recent course teaching experience in the course they are placing for, and they must be limited to making a decision of either into or out of that course. Since Course 11A no longer existed as a course at Cardinal College, and at the time of collecting this data, hadn't for five years, it may be problematic if Course 11A placement is decided first before deciding Course 11. Yet, it is also logical for Theresa to have continued to use her Course 11A knowledge, as she was tasked with the challenge of figuring out what to do with these students when neither placement option was ideal. She turned to what she knew of teaching the College's particular blend of first-generation and international college students, and focused on the program's goals for teaching and learning in these courses. Yet, since the use of her Course 11A knowledge is both potentially valuable and potentially problematic given Smith's expert-reader model, I want to continue to consider both of these viewpoints. Next, I discuss one more example and contrast it with student 1011-03-02, discussed earlier.

With student 1011-09-06 in Redo Session 2, Theresa again, like she did with 1011-03-02 earlier, made a comment on teaching. However, she ignored her Course 11A knowledge and pushed the student down. She read the student's piece and noted that the student presents a misreading of Graff as "all sports, no education" in the "first four words" of her response. Yet, later the student "gets down to the point where she's able to talk about evidence and stuff" (Redo Session 2). Theresa continued, reading a few more lines of the student's response aloud, and said:

This is another paper that is. Where the student is in no way fully prepared for college literacy. But is also probably enough of a decent reader that a four-credit course [Course 11A] would serve just as well as the six [Course 10] but I can't put him or her in because the misreading of Graff is egregious. (1011-09-06, Redo Session 2)

Having decided that the student would do best in Course 11A, which is not an option, Theresa said:

When they come down and actually apply themselves to Pluto, they actually start to get somewhere. And if this paper were revised and put through some kind of a process, they would probably be alright. But I can't say in. I have to say not in. (1011-09-06, Redo Session 2)

Here she reminds herself of the procedure she should be following—only deciding if the student is in Course 11 or not.

Because she pushed this student down, the student's response goes on to be read by Adele. After noting the same issues with reading at the front, Adele pushed it back up because she also noted where the student gets to later in her piece:

This [end point] is actually really good. But they sort of miss it up at the top. But I think this could be a tweener person. Definitely I mean cause what they do at the end here is pretty good. But they just. They have to make a stronger connection with the first reading. (Adele, 1011-09-06, Redo Session 2)

Here Adele also said that revision would fix this student's issue. Therefore, in the tweener conversation, Theresa and Adele decided quickly together that the student should be placed in Course 11. Thinking about how the course emphasizes writing as a process

helped decide this student's placement. But the far more interesting aspect is that even though Theresa made the revision observation alone, she did not decide Course 11 for this student on her own. She used the expert-reader reading process, and it turned out the Course 10 teacher pushed the student up as well.

It may appear on the surface that the extra reading was a waste of time, as both readers came to the same assessment. But as Smith's model emphasized, and as my analysis of the criteria of placement at Cardinal College has shown, this shared decision is not assured. I argue that while Theresa's prior knowledge is essential to the weighing of the reconfiguration of Cardinal College's course boundaries, it is only half of the equation. If Course 11A students have attributes that place them between the two existing courses, the program has to be able to have conversations about where to redraw the line, and that only happens when experts from both existing courses get to weigh in on problematic students.

Recommending a particular teacher or teachers. In addition to the three types of readers' comments already discussed, I also observed Theresa, the writing program director, make a fourth kind of comment that she came to refer to negatively as "reading like an administrator" as she made her placement decisions (1010-01-11, Redo Session 1). Reading in this way, Theresa would frequently match students to teachers as part of her placement decision, noting attributes of a particular student that would make her recommend a particular teacher as the best possible placement. After making a placement decision, Theresa would record a suggested teacher or teachers for a student, if she had one. This note went with the essays to the advising office for their use in building the student's schedule.

The use of this knowledge is an interesting facet of placement at Cardinal College to consider, as it is logical that the strengths and weaknesses of particular students could find their match in the strengths and weaknesses of particular teachers. Yet, it is also potentially problematic as choosing the teacher could be seen as muddying the water, making it difficult to focus on the student's fit to the course of expertise rather than the preferred teacher of the course. After providing some sense of how frequently these recommendations are made, I offer an analysis of key examples from each session and weigh the value of teacher recommendations on how they relate to privileged course knowledge against their potential to disrupt the expert-reader model on which Cardinal College operates.

Original sessions. For the 26 participating students in the original placement session, Theresa recommended a teacher for 9, or 34.62%. But the writing program's records, which include teacher recommendations, allowed me to count how frequently this occurred in all of the original placement sessions, including the ESL placement session held at the end of the previous spring semester a few weeks before summer placement sessions started. Out of 224 students who sat for the placement in these sessions, 41, or 18.30%, had a teacher recommendation on their essay.

And these recommendations can be seen as potentially helpful. Take for example student 1011-02-03, where Theresa recommended either Odysseus or Terence because the student was off assignment and needed someone who wouldn't "put up" with that (Original Session 6). Or, for example, student 1011-03-03, who Theresa saw as having a set of quirky issues combining complex thinking with significant grammatical errors, and she felt she understands that well, so she recommended herself (Original Session 7).

These examples make it clear that choosing a teacher could be a valuable way of matching students to particular teacher's strengths.

But they are also problematic because the concern with which teacher can distract from the placement decision. For example, as she discussed student 1010-04-08 during Original Session 5, Theresa noted several mixed issues with the student. First, he had one paragraph but it had moves inside of it, and she talked about how he had a move that was meant to "simplify" the text to something he would be able to deal with in his writing rather than to reduce it, describing this as "not a Course 10 move" (Original Session 5). As she continued, she then noticed that the student "[jumped] off" the assignment and went into an essay about his high school, which caused her to decide he needed to go down to Course 10. However, she said immediately after that "he shows some signs of being a tweener" and "I don't care what section he ends up in [Course 10] or [Course 11]. I would like to have him in my class," feeling she could handle his particular issues. Here she pushed the student down to Course 10, but she indicated that she was far less concerned with which course, and more so with the teacher placement. While there were no other moments in the transcripts that suggested that the choice of teacher took precedence, and even this moment was only a suggestion, as she made her placement decision first, we have to remember that think-alouds are not a full reporting of thoughts (Ericsson & Simon, 1980, 1984). We can't be one hundred percent sure which takes precedence, the course placement or the teacher placement, although the data here suggests it is the course placement decision. Yet, there are other indications that the choice of teacher really matters in ways that could potentially muddy the expert-reader model as we will see more clearly going forward.

Redo sessions. In the redo placement sessions, the readers—while reading essays blind of marks, names, or previous decisions—knew these sessions were not “real” and that these students had already been placed. As such, the frequency of teacher recommendations dropped dramatically. In fact, only once did Theresa specify an exact teacher (1010-06-08); however, on three other students, she makes statements about what kind of instructor she would choose. These three comments, when taken together with a comment Theresa made upon finishing her reading of student 1010-04-05, help us understand that Theresa chooses teachers in direct relation to the determination of a student as belonging in the old Course 11A or not.

On student 1010-05-02, Theresa said the student “needs [Course 11A] with the right teacher” (Redo Session 2). As far as what she meant by the “right teacher,” this single moment didn’t specify, but when we look at what Theresa said about student 1011-03-01, we discover she made the same statement but added “which would be a [Course 11A] instructor” (Redo Session 2). This specification that the right instructor is an old Course 11A instructor clarifies why she often chooses the instructors she does. In fact, while she only said, “I would be careful where I put that person” on student 1011-09-14, I take her to mean by “careful,” that she would choose a Course 11A teacher (Redo Session 1). The point is that even though she chooses a teacher to address particular student weaknesses, these weaknesses are usually those characteristic of Course 11A students.

Finally, after finishing her discussion of student 1010-04-05, another student who she felt showed the need for Course 11A, Theresa said, “And that’s why my Course 11 sections look like they do...because I throw all these people in my class...But that’s not a

structural solution to the problem.” She continued, “You cannot have a curriculum that’s running basically because one person’s reading the placements and recognizing that there needs to be a different course and therefore creating it inside the system” (1010-04-05, Redo Session 2).

Looking closely at moments when a teacher is recommended, and at Theresa’s own admission on what guides these choices, it is clear that the majority of these comments recommending a teacher relate to a concern with putting Course 11A students with those teachers who have experience teaching old Course 11A students. Even in instances from the original placement sessions where non Course 11A instructors were recommended (as in the case of a few students I’ve discussed above), these non-Course 11A instructors were usually recommended for students who were determined to be unproblematic, and who would have no trouble in a standard Course 11.

Findings on Readers’ Criteria Being Teaching Specific

The analysis of Cardinal College’s placement process through the think-aloud protocols of its readers showed instances of what Smith (1992/2009) called privileged course knowledge, and two categories of comments were shown. Placement readers use their knowledge of what is typical in their course of expertise as they consider where students may struggle or excel, and as they think about how they will teach students with specific issues. The identification of these instances shows that Cardinal College’s placement readers indeed use teaching specific course knowledge to place students, and such knowledge is particularly helpful when readers must determine the placement of problematic students.

However, potential problems were also shown when readers noted students who did not fit neatly into their courses. At times in these moments, Theresa, the Course 11 reader, relied on her expertise from an old course option, and she attempted to solve problematic placement decisions on her own by using her old expertise or by recommending other Course 11A teachers rather than by pushing students down. Her desire was to tailor placement to ensure that a particular subpopulation of student needs were met without unnecessarily subjecting these students to the burden of additional credits and the resulting consequences of that burden.

Tailoring placement decisions to students' needs is, in its intent, exactly what placement does, and Smith's (1992/2009) sense of placement *adequacy* is at the heart of Theresa's decision making, as is the consideration of the consequences of placement decisions (Messick 1989b), particularly for students who must take on the larger credit load. In these ways, Theresa's intent is precisely aligned with all the desired outcomes of adequate placement decisions. However, the focus on an old course and the choice of a teacher are potentially problematic, and we can see this by thinking about the situation hypothetically. What happens in the reader's mind when the placement decision and choice of a teacher for the student conflict? For instance, what if there is no section of Course 11 open or being taught by an old Course 11A teacher? Would that have changed the reader's placement decision, potentially pushing the student down to Course 10?

In short, in a writing program that does not include Course 11A as an official registered class, there is no way to ensure a Course 11A student receives a Course 11A instructor, so this distinction and decision at placement can negatively impact both the reliability and adequacy of placement decisions. Furthermore, it is not really possible to

claim to use Smith's (1992/2009, 1993) expert-reader model when there is no Course 11A on which to be an expert. A reader who is focused on Course 11A will invariably find themselves, as Theresa does, stuck trying to weigh possible placements against each other, rather than making a decision about one course. And, since course expertise is built on experiences with students, solving the problem through teacher choice means that teachers in the program will have two different sets of experiences with students in Course 11. The result may be that Course 11 teachers will not see certain student issues as part of what is normal for their course, and the placement of the problematic student population with only certain teachers will prevent the whole of the writing program and its courses from adapting to the loss of the Course 11A option but the retention of its student population.

Are Expert-Readers' Placement Criteria Linked to Course Criteria?

In order to consider if the placement criteria, sketched earlier, were course related, I looked for links between the placement criteria and teachers' course criteria. To uncover the teachers' course criteria, surveys of adequacy and think-aloud protocols were collected as teachers read diagnostic essays and checked placement for each section of their course during the first two weeks of the semester in August of 2013 (see Appendix L for a copy of the survey, which asked teachers to rank students as prototypical of their course, as slightly above or below it, or as having been improperly placed). This section discusses data from the think aloud transcripts and surveys of each of the seven participating teachers. It is necessary then to begin with a review of some information from earlier chapters that will contextualize this data and its analysis.

I start by offering an overview of the diagnostic reading sessions, more specifically a summary of the numbers of participants and decisions made, as well as a description of the three diagnostic prompts teachers administered to their students. Next, I give a description of the reading and writing criteria that emerged from coding and analysis of these transcripts overall, before providing important notes on how I coded and grouped the data for analysis that will be relevant to how I've organized my findings in this section. Once this context is established, I proceed to present the discussion of the data and results of my analysis.

Overview of Diagnostic Reading Sessions

Seven participating teachers completed think-aloud protocols and a survey of adequacy for each section of their courses as they read the 109 diagnostic essays submitted by participating students. An additional nine participating students failed to submit a diagnostic essay. Three of the teachers—Daria, Adele, and Hannah—taught Course 10. Daria taught sections 01 and 06. Adele taught sections 02 and 07, and Hannah taught section 05. Three teachers—Terence, Odysseus, and Janet—taught Course 11. Terence taught section 02. Odysseus taught section 03, and Janet taught sections 05 and 09. One teacher, Theresa, had one section of Course 10, section 04, and two sections of Course 11, section 07 and 08. There were a total of 55 Course 10 student essays read in the diagnostic sessions with Daria having a total of 25, Adele 10, Hannah nine, and Theresa 11. There were a total of 54 Course 11 student essays read in the diagnostic sessions with Terence having a total of eight, Odysseus 12, Janet 20, and Theresa 14. There were no participating students who the teachers felt were improperly placed. Yet, 9 participating students in Course 10 were ranked as slightly below and 12 were slightly

above. The remaining 34 were ranked as prototypical. In Course 11, two students were ranked as slightly below, 11 as slightly above, and 41 prototypical.

A Description of the Three Diagnostic Prompts Teachers Used

As teachers in the program are always asked to administer a diagnostic essay to their students during the first two weeks of the course to recheck placement, Theresa stacked copies of diagnostic prompts in the faculty copy room so that teachers in the program could use one of her diagnostics if they did not have their own. The prompt she shared this semester, used by all teachers in the study except two, asked students to read an excerpt from Gerald Graff and Cathy Birkenstein's *They Say, I Say* which discusses academic writing as a conversation or debate, as well as a piece by Stephen Ambrose that makes an argument about the reputation of General Custer. The directions asked students to "write an essay" that "[explains] Ambrose's position in the debate about the reputation of George Armstrong Custer." Citing a phrase from the Graff and Birkenstein piece, it also asked students to consider "what [they] might need to know to 'put [their] oar in.' What might [they] or others say in response to this debate?"

Adele had used this diagnostic before, and since she had been using it for some time, she had rewritten the directions to make it clearer to the student that they were being asked to use some understanding from Graff and Birkenstein's text to think about what Ambrose says and does as a writer. Therefore, while her students wrote on the same essays, their diagnostic prompt asked them to "write an essay in which [they] used [their] understanding of Graff and Birkenstein's perspective on academic writing to critically analyze what Ambrose is doing as a writer." Adele's prompt also asked students to

“consider what [they] might need to do to ‘put [their] own oar in,’ or enter the conversation about General Custer.”

Odysseus chose to use his own prompt and readings. Specifically, his prompt provided students with the full text of a short essay by Aldo Leopold called “Thinking Like a Mountain.” Students were also given a quote from Leopold’s “The Role of Wildlife in a Liberal Education.” They were then given the following directions:

Your assignment is to write an essay in which you apply this quotation to a reading of “Thinking Like a Mountain.” In what sense is the latter essay an example of “liberal education in wildlife”? How, in other words, does this essay enact the goals, practices, and ideals that Leopold expresses in the above quotation?

Review of the Reading and Writing Criteria

As all think-aloud transcripts were coded, both placement and course transcripts, 14 criteria of reading or writing that placement readers and course teachers discussed on students’ essays emerged in total. Table 2, presented earlier in the chapter, lists and describes these criteria. All of the 14 criteria were discussed in relation to diagnostic essays. It is important to recall that these criteria were developed directly from the language of the placement readers and teachers, and thus some of the terms and understandings are particular to the context of Cardinal College. See the earlier section describing the reading and writing criteria for clarification if necessary.

Reminders on Coding and Analysis

As I began the discussion of the placement transcripts earlier in the chapter, I reviewed some important aspects of coding the data. The same coding and notations are

relevant to these diagnostic think-aloud protocols as was to the placement ones. Specifically, teachers discussed many of the 14 criteria as existing on a continuum of ability with a comment often having a negative or positive connotation, and readers often commented on aspects of the criteria that could be considered subcategories. In addition to noting these aspects, I also counted the number of students on which each criterion was discussed by each teacher, as well as if the discussion of a criterion was negative, positive, or mixed overall. These counts allowed me to see which criteria were more or less important to teachers when they determined how students fit into their course and they allowed me to characterize the discussion of those criteria. More will be said on these issues as I discuss and quote specific comments from the data as they are relevant to my findings.

Finally, after finishing coding and moving to analysis, I analyzed the students overall, but I also separated and looked at students who were ranked in similar ways by course teachers—that is, I compared all prototypical students of Course 10, all students slightly above Course 10, all those slightly below Course 10, and the same three groups for Course 11 (see Appendix T for a table listing all 109 course students and their ranking on the survey of adequacy). I also compared students ranked slightly high for Course 10 with those ranked slightly low for Course 11, as these students were the participating students who were closest to the old Course 11A designation noted during placement. In what follows, I discuss the results of my analysis of the course transcripts.

Overview of Course Think-Aloud Analysis Results

Going forward, I present the results of my analysis of the course diagnostic transcripts. I begin by providing the results of the adequacy surveys counting all students

in the courses who submitted a diagnostic essay. Next, I present the course criteria overall, including the combined total number of students for which each criterion was discussed by at least one teacher. While the rank of criteria that results by totaling the number of students for which each criterion was discussed is skewed so that teachers with more participating students end up having their criteria play a more significant role in the overall total, I offer it as a starting point in seeing how these criteria are ranked. Next, I detail the Course 11 criteria and the Course 10 criteria. Since no participating students were designated as improperly placed by course teachers, and since so few in each course were flagged as even slightly atypical, I chose to fold the discussion of slightly below and slightly above students into my analysis of the prototypical course criteria, rather than separate them out into their own section. In short, I use the slightly atypical students of each course inside a discussion of that course to think about which criteria teachers find most concerning or most impressive and hypothesize which criteria may represent the edge of these courses.

Specifically, with Course 11, I show that the Course 11 teachers' criteria differ significantly from each other and from the placement criteria, and I discuss how these differences are likely caused by the population of students that individual teachers have experienced, both in the past and presently. With Course 10, I show the consistency in teacher's criteria, and I argue that this consistency can be potentially useful for helping the writing program understand the most concerning habits of reading, a necessary understanding if the program is to redistribute the old Course 11A student population into the two existing courses.

Adequacy Survey

I first looked at the teacher's survey of adequacy results on the total student population. While Smith (1992/2009) discusses how this measure is potentially problematic due to teacher's reluctance to move students, it is still an important starting point and one set of data that can triangulate with others. When asked to rank their students on a survey placing students as prototypical of their course, slightly above, slightly below, belonging in the higher course, or belonging in the lower course (if applicable), course teachers only indicated that three out of the 175 students, or 1.71% who submitted a diagnostic were improperly placed. This number shows an overall satisfaction by teachers with the placement decision. In fact, only two teachers who were both Course 11 teachers, Odysseus and Theresa, indicated there were students improperly placed in their courses, and it is a testament to Theresa that she was, in essence, potentially challenging her own decision on two students.¹ But again, since teachers are typically reluctant to move students, as Smith said, it is worthwhile looking at how many students were flagged as slightly below or slightly above, as it is possible that teachers could have actually felt that some of these students were improperly placed as well.

For Course 10, there were 55 students who were prototypical, 17 students who were felt to be slightly below, and 15 who were felt to be slightly above out of the total 88 students who submitted a diagnostic essay. Since there is no course option below Course 10, none of these students could be moved anywhere regardless of if the teacher felt they really were below the course or not. It is possible that some or all of the 15 students who were felt to be slightly above could really have been potential candidates

¹ I say potentially here because neither of these students was participating in the study, so I cannot check the original file or record of either to see when their decision was made. It is possible that they were students who graduated up from Course 10, but who still displayed some weaknesses.

for moving up. For Course 11, there were 58 prototypical, 13 slightly below, 14 slightly above, and the aforementioned 3 who it was felt should be moved to Course 10. If we add together all the students who could be moved—Course 10 students suggested as slightly above and Course 11 ones as slightly below, slightly above, or should have been moved—we get 45 out of 175, or 25.71% as potentially misplaced, although it seems highly unlikely that teachers would feel all of these students were improperly placed, as all courses have some variation. What these two numbers do is leave us with a low end measure, the 1.71% who were suggested for moving, and a high end, the unlikely worst case scenario of 25.71% misplaced.

Course Criteria Overall

To get a sense of which criteria mattered most to teachers in the courses, I counted how many of the total 109 course students had a particular criterion mentioned in the discussion of their diagnostic essay. As I noted earlier, certain teachers had more students participating than others. Therefore, this count, which is a total, favors the criteria of those teachers who discussed more students. However, I present the total as a starting point for discussion. Table 11 presents the total number of course students discussed for each of the 14 reading and writing criteria.

When the discussion by all teachers on all 109 students in the diagnostic think-aloud sessions is accounted for, the most important criteria for determining a student's fit to either course are the students' quality of reading, how they work with the text, their understanding of the assigned task, and their knowledge of paragraphing or essay structure. Several other criteria are important overall, but to a lesser degree. On the face

then, these criteria do not seem so far from the ones important for placement, as the same main criteria are discussed during the courses as well.

Table 11

Total Number of Course Students Discussed for Each Course Criterion

# of Students	% of Students	Name of Criterion
90/109	82.57%	Reading issues/quality
84/109	77.06%	Working with text
70/109	64.22%	Assignment task
61/109	55.96%	Essay/ paragraphing knowledge
44/109	40.37%	Framing/moves
41/109	37.61%	Grammar
28/109	25.69%	Length/amount of writing
27/109	24.77%	Use of required texts
26/109	23.85%	Language
26/109	23.85%	Academic rhetorical conventions
25/109	22.94%	Thinking
14/109	12.84%	Thesis/project
11/109	10.09%	Print conventions
3/109	2.75%	Writing process

Course 11 Criteria

Now that I have described the course criteria overall, I want to move to detail the criteria of Course 11 more specifically and compare the criteria of the course to the placement criteria related to that course. Specifically, I argue Course 11 teachers' criteria differ significantly from each other and for some teachers, from the placement criteria. Given the findings of previous sections, I believe these differences could be caused by the population of students that individual teachers have experienced, both in the past and presently.

It is my theory that if a Course 11 teacher's past experiences with teaching were primarily with unproblematic Course 11 students, their assessment of how well students fit Course 11 shows less concern for reading, both as an individual criterion and one

impacting other criteria. But, if a Course 11 teacher has had more experiences teaching Course 11A students, when the course existed, their assessment of how well students fit Course 11 shows more concern for reading, both as an individual criterion and one impacting other criteria. Additionally, since Course 11A knowledge has been shown to influence placement, particularly with the recommendation of old Course 11A teachers for old Course 11A students, it is possible that teachers with primarily Course 11 experience have not seen enough of a change in student population and have not had enough direct experiences with students that have more severe reading issues for their sense of what is normal for the course to have changed, or for their sense of what they should do to address these issues to change.

Looking at the think-aloud and interview transcripts, I first consider two Course 11 teachers, Theresa and Odysseus, who had extensive experience teaching Course 11A, and I show that, for both, reading is the prime criterion and many others concerns are related to the concern with reading. I contrast these teachers' concerns with reading against the criteria of Terence and Janet, two Course 11 teachers who taught Course 11A far less, if at all. Looking at the discussion these two teachers have about reading, and related criteria, I show how they do not have the same concerns about reading.

Theresa's criteria. Since Theresa is the Course 11 placement reader in this study, I start with her criteria. Establishing the connection of her reading of diagnostics here to the placement criterion of reading should be fairly straightforward. First, let me present, in Table 12, the total count of criteria discussed for the 14 participating students in her two sections of Course 11.

Table 12

Total Number of Course 11 Students Discussed by Theresa for Each Course Criterion

# of Students	% of Students	Name of Criterion
12/14	85.71%	Reading issues/quality
11/14	78.57%	Working with text
5/14	35.71%	Essay/ paragraphing knowledge
5/14	35.71%	Framing/moves
4/14	28.57%	Grammar
4/14	28.57%	Academic rhetorical conventions
3/14	21.43%	Assignment task
3/14	21.43%	Use of required texts
2/14	14.29%	Language
2/14	14.29%	Length/amount of writing
2/14	14.29%	Thinking
1/14	7.14%	Thesis/project
1/14	7.14%	Writing process
0/14	0%	Print conventions

Thinking back to both Theresa's unproblematic and problematic Course 11 criteria, this frequency count is not all that dissimilar. During placement, Theresa talked about reading on almost all Course 11 students' essays. In addition, she was concerned with the students understanding of the assignment, as well as their work with the text, as they related to reading. While many of those same issues top the count here, the most shocking difference is the drop in how frequently Theresa discusses the assignment task. However, while these numbers seem to suggest a difference, a look at the qualitative side of this data reveals that the differences are not that stark, and reading still is the number one issue for Theresa.

The primacy of students' reading. Reading is still the main attribute Theresa is assessing on students' diagnostic essays. Of the 14 students she assessed here, 12 were commented on for reading. Thus, she is still primarily concerned with students reading. And just like at placement, she tried to determine if a student's misreading would be

serious or if they would be able to work a misreading out by being taught close reading. For example, with student 1011-08-07, Theresa noticed that the student tried to make a reading connecting the texts, but the student used imprecise language to make it, saying “this” rather than naming what she sees. In addition, she lacked direct work explaining the text and instead had “quotations...embedded in her sentences.” But Theresa determined that “If you put her head back in the text, she will be fine” because “She’s not all that far off” (Diagnostic Session). Theresa made similar comments with six other students’ pieces, and she implied with many, as she does here, that her concerns with the students’ reading issues will be addressed by the course.

Work with text. Given Theresa’s concern with students’ abilities to use close work with text to solve their reading issues, we see already that, just as at placement, the way students work with text is intimately connected to their reading for Theresa. Additionally though, some students in Theresa’s section of Course 11 share some similarities with the Course 10 placement students, particularly in regards to copying. For example, with student 1011-08-04, Theresa said, the student was “largely repeating back what is in the passage. Just repeating back whole sentences” and she added “quoting and not quoting” at times (Diagnostic Session). How this student worked with text resembled issues Course 10 students had. This student, as a point of difference, seemed to recognize she was quoting at times, but at other times she did not. There is another student who displayed some less severe issues failing to mark quotation boundaries, student 1011-08-05. Both of the above students were ESL students, and Theresa considered that their issues were likely a result of the on-demand nature of the task, but additionally, “cultural” (1011-08-04, Diagnostic Session). Theresa also knew, as this was week two or three of

the course, that both students had already sought out help at the writing center, and as such she was less concerned.

Essay and paragraphing knowledge weighed against reading. Despite not discussing this criterion as frequently as she did during placement—only five of the 14 students here have a consideration of their essay and paragraphing knowledge—Theresa was still concerned, as she was at placement, with weighing essay and paragraphing knowledge against a student’s reading ability. For example, with student 1011-07-02, Theresa described the student’s issues with reading as mostly “retelling” the texts and then “[giving] his opinion” on Custer, only slightly trying to connect the two passages, but Theresa felt, despite this misreading that “he’s got more of the essay structural things” and the reading was a “higher level” misreading (Diagnostic Session). This kind of comment is characteristic of Theresa’s feelings on Course 11 students’ knowledge of essay and paragraphing knowledge. If the student possessed some knowledge in this area, her concerns about their reading were quieted some.

Assignment task discussed less than at placement. The decrease in Theresa’s comments on how students understood and addressed the assignment task has a lot to do with her sense that the diagnostic prompt she gave students was the problem. She realized that she had copied an earlier draft of the prompt that didn’t “cue [students] to look at Ambrose and Utley” (1011-07-03, Diagnostic Session). In fact, when she praised student 1011-07-03’s reading of Graff and Birkenstein, but then hit a moment in the student’s text where she noticed the beginning of a misreading as the student turns to connect Ambrose to Graff and Birkenstein, she said the misreading was “Almost a misreading that the assignment invites in a way” because the assignment didn’t point students to look

at Ambrose's work as a writer with Utley (Diagnostic Session). You can see this same concern with her assignment in the comments she makes on two other students. With student 1011-08-10, for instance, Theresa noted a student who connected the texts on a kind of misreading, but she said, "I can understand because of the passage [meaning prompt] why they went with the Custer side and not the Ambrose and Utley" (Diagnostic Session). In short, Theresa believes that the poor wording of the diagnostic prompt has caused some of the students' issues, and she adjusted her reading accordingly. This problematic view of her prompt is likely what accounts for why she discussed this criterion less frequently during the diagnostic sessions, as she has discounted that as a fair factor for these students.

Theresa is a Course 11A teacher. If it is not already apparent from the description of Theresa's concerns so far, Theresa reads diagnostic essays, as she does placement essays, with a focus on the problematic Course 11A student's attributes. This is a fact she admitted herself in a comment she made after finishing her think aloud on student 1011-07-04. After choosing to rank the student as slightly above on the survey she was completing, Theresa said by way of clarification, "When I think of my prototype, I am looking at what I would call these. Not Course 11 but old Course 11A. My prototype is a person who is not fully prepared but not grossly underprepared." She continued, "Somebody else's prototype might be different, but I know what my class has in it because I have been putting people in it all summer" (1011-07-04, Diagnostic Session). Here she admits, as she did during placement, that she has been placing students into her particular sections of the course when she had concerns about their

being old Course 11A students, and she expects to find that her students show more concerning issues than the majority of students in other sections.

Odysseus's criteria. I begin again with an overall count of Odysseus's diagnostic criteria as a starting point for discussion. Table 13 displays the total number of students discussed by Odysseus for each criterion. While reading does not top Odysseus's list, work with text does, and as we have seen in the past, work with the text can be a criterion that is linked to reading. Let's investigate the think-aloud transcripts more closely to uncover patterns in Odysseus's discussion of Course 11 students, particularly in regards to reading.

Table 13

Total Number of Course 11 Students Discussed by Odysseus for Each Course Criterion

# of Students	% of Students	Name of Criterion
12/12	100%	Working with text
11/12	91.67%	Essay/ paragraphing knowledge
10/12	83.33%	Reading issues/quality
8/12	66.67%	Assignment task
8/12	66.67%	Language
8/12	66.67%	Use of required texts
6/12	50%	Academic rhetorical conventions
5/12	41.67%	Print conventions
3/12	25%	Grammar
3/12	25%	Framing/moves
3/12	25%	Length/amount of writing
2/12	16.67%	Thinking
2/12	16.67%	Thesis/project
1/12	8.33%	Writing process

A students' lack of work with the text calls their reading into question. The most frequently discussed criterion by Odysseus is work with the text. He discussed this criterion on all of the students' essays, sometimes positive, sometimes negative, and sometimes as a mix of the two. The troublesome students discussed by Odysseus for this

issue are particularly important as these students appeared, on the surface, to have some sense of what they read, but Odysseus was hesitant to trust that their reading was not without its problems because these students had little or no direct work with the text.

Before we look at the students who did not work with the text, let's first consider student 1011-03-10, who Odysseus felt did do some work with the text. Odysseus said the student "does want to do some work with the lens text and actually does more." He continued, "This is somebody who...would be a little higher than the norm, slightly above, but still doesn't quite get what the lens text means and how it applies. He keeps saying it applies." As Odysseus reread a moment from the student's piece, he said, "Yes indeed he did, but how?" in response to the student's claim that "Leopold...provided an example of liberal education in wildlife." He concluded, "The student doesn't know how to say that" (Diagnostic Session).

What Odysseus says here presents a good contrast to what he notes later with worrisome pieces. Here, he says the student claims an idea he doesn't show, but he believes the student sees some ideas because he does work with the text in other places. Alternatively, he doubts the reading of other students when he doesn't see these students work with text or effective work with text. For example, with student 1011-03-01, Odysseus said, "And they're putting in quotes cause they know they are supposed to do that. And they know they are supposed to talk about quotes, but the quotes are just sort of carrying them wherever the quotes will carry them," noting later that the student turned to cliché with "pretty soon he's going to talk about the circle of life" (Diagnostic Session). Here we see that Odysseus distinguishes between knowing you should quote, and knowing to do close work with the text to explain a quote.

And on student 1011-03-04, Odysseus said, in response to a moment of the student's piece that indirectly used an idea from one text left unmentioned "That's something that is referenced in the other essay...so sort of in the back of this writer's mind must have been the other thing that was read," but "it's not specifically referred to. It's not quoted." After reading more, Odysseus concluded, "Doesn't know really how to work with the text and gets lost in the complexity of the ideas of the materials," signaling that the student's issue in reading, the issue in getting "lost in the complexity," is in part due to his lack of textual work (Diagnostic Session). Odysseus, like Theresa and like the placement reading for Course 11, clearly sees reading as a key issue of the course.

For Odysseus, Course 11 students typically have either reading or essay knowledge issues. Many times as Odysseus was discussing students in the course, he characterized students as falling into one of two camps: students who struggled to read or who struggled to represent their ideas in the structure and form of an essay. Both sides can be seen in what Odysseus says about student 1011-03-02, who struggled with the reading end. Having decided the student has "the structure of an academic essay," and "understands the essay a little bit better" than he understands how to "work with the texts," Odysseus called the student prototypical (Diagnostic Session). Having just designated an earlier student as prototypical for the opposite issue of understanding the reading but not writing an essay, Odysseus added:

I see both sides of that coin. Somebody who knows how to do something with text. Somebody who knows how to structure an essay. It's pretty rare, as a Course 11 person, that they know how to do both...especially with difficult materials.
(Diagnostic Session)

So prevalent was this distinction, that a large number of Odysseus's students can be classified, from his think-aloud descriptions, as following on one or the other side. Six of the 12 students had some essay knowledge, with at least a three paragraph essay, but they struggled to read the texts or to work adequately enough with them. Two of the 12 students had stronger reading, but lacked essay knowledge. For example, one student wrote a one paragraph, "stream of consciousness" response (1011-03-06, Diagnostic Session). One student, ranked as slightly above, had some understanding on both sides. The remaining three students had trouble with how they work with the text or failed to work with the text, which called their reading into question. From his discussion of these two criteria alone, it seems in Odysseus's experience, reading issues are prevalent in the course, as even though he splits students into these groups, the larger group has trouble with reading. Since Odysseus has experience teaching the old Course 11A, and since problematic students were often recommended for his sections, we see reading is a key attribute he assesses as he reads essays related to his diagnostic.

Odysseus has designed his prompt to expose reading problems. During the diagnostic reading sessions, directly after he finished discussing student 1011-03-12, who Odysseus concluded "[seems] to understand the connection" between the two texts by Leopold, but does not really, Odysseus talked about the design of his prompt. He said:

I really like my diagnostic. Because that's what happens. They fall back on the easier text, and it's a way that I can see who can read at what level. Cause when they have recourse almost entirely to "Thinking Like a Mountain," it shows that that's where they are as readers. They can read a fairly straightforward text....But then when it comes to that other text which asks them to think about the nature of

education, how we are educated, whatever else, and then to apply that to the one, they can't do that by and large. And that gives me a reading on where, on what I need to do with them. Which is obviously to understand how to use more difficult text. (Odysseus, Diagnostic Session)

By his own admission, the goal for administering this prompt is to expose students who need to learn to work with the text. And since for Odysseus working with the text is related to reading, we see that what Odysseus learns about the students reading is essential to his assessment, and his course planning.

Additionally though, Odysseus makes similar comments about the placement prompt, and what he says further emphasizes the importance of reading as a criterion. When asked to discuss, during the interview, what he believes the placement prompt asks students to do and how he feels about it, he said,

I think this is a good prompt because it does ask them to take a difficult piece, namely Graff's, which is way beyond anything they've even thought about when they come in here...and then have them map that on to something which is within their comprehension. (Odysseus Interview)

While he went on to stress that the placement prompt matched what he does in the class, he also expressed a concern about the Pluto piece "being popular journalistic writing" and therefore "so darn easy" that "a student who ignores Graff all together, as many of them do, can write a fairly coherent essay" (Odysseus Interview). Odysseus was concerned that on that such a student could be placed into Course 11 but go on to "flounder when they encounter tough stuff" (Odysseus Interview). So in essence, seeing

what Odysseus said about his diagnostic and about the placement prompt shows the importance he places on reading as a key deciding criterion of placement.

For Odysseus, assessing their reading is key to knowing if a student can do academic work or not, as he believes many students may be able to fashion a piece that uses the forms and structures of academic writing, when tasked with relatively simple readings and familiar concepts, but how they react to difficult texts, unfamiliar concepts, and academic reading tasks better reveals if and where they will struggle in the course or not.

Terence's criteria. While Theresa and Odysseus have extensive Course 11A experience, the same is not true of the other Course 11 teachers. Terence may have taught a section or two of Course 11A over her years at the College, but she much more frequently teaches Course 11 or Course 11 Honors. As such, her think-alouds and the resulting criteria show a different relationship to reading. Table 14 displays the total number of students discussed by Terence for each criterion. What is most striking here is that Terence discusses reading on 100% of these student essays, but examining what she says about reading closely shows a key difference to the importance she places on this criterion for making fit-to-course determinations.

Problems with reading are not as concerning to Terence. While Terence discussed reading on every student essay, what she said about the students' reading was usually not a cause for her concern. Reading, it seems, is not a factor in determining a student's fit to the course for Terence, although admittedly it was harder to discern from Terence's data what criteria might be concerning since, she flagged all but one of her students as prototypical on the survey of adequacy when other teachers averaged 5.27

Table 14

Total Number of Course 11 Students Discussed by Terence for Each Course Criterion

# of Students	% of Students	Name of Criterion
8/8	100%	Reading issues/quality
7/8	87.5%	Assignment task
7/8	87.5%	Essay/ paragraphing knowledge
7/8	87.5%	Grammar
6/8	75%	Working with text
5/8	62.5%	Language
4/8	50%	Framing/moves
3/8	37.5%	Academic rhetorical conventions
3/8	37.5%	Thinking
2/8	25%	Use of required texts
2/8	25%	Thesis/project
1/8	12.5%	Writing process
0/8	0%	Length/amount of writing
0/8	0%	Print conventions

students in the slightly above and slightly below categories combined. Yet her students did show variations in their reading abilities, with three commented on for strengths, two with mixed assessments, and three with negative comments. Two of the students with negative comments (1011-02-07 and 1011-02-03) seem, in comparison to others, to have more severe reading issues. But, other than noting this fact, Terence was not alarmed in terms of their preparation for the course.

On student 1011-02-07, Terence noted that he had paragraphs and started with the *They Say, I Say* texts, which Terence considered “a good idea because that’s what we sort of asked them to look at the Custer piece through,” but as she continued reading, Terence noticed the student said that Custer was “a great example of thinking critically and intellectually. The passage explains that in battles Custer could time his moves, or fall back with almost flawless execution every time” (Diagnostic Session). In response, Terence noted that the student had misunderstood the prompt and thought that he should

be “worried about Custer being the critical thinker” and had therefore missed the “two scholars” and their arguments about Custer, seeing only “one position” (Diagnostic Session). As she finished her reading, Terence concluded that

This kid is fine. He didn’t completely understand the reading or the prompt. But he can put a sentence together. He knows that you are supposed to have paragraphs. He introduced the topic. He did his best to conclude it. Yeah, I think we just need to work on like reading and understanding. (Diagnostic Session)

For Terence, the student’s reading issues are balanced by his other knowledge. Like many of the other teachers, she seems to weigh these criteria against each other.

Yet, there are other indications that reading, particularly difficulty with the kind of reading called for by the prompt, is not surprising to Terence, as she feels those factors aren’t what she should focus on when deciding if a student belongs in her course. As student 1011-02-03 discussed the details of Ambrose’s piece on Custer, Terence realized the student had “completely misunderstood the passage” and that the student missed that the text was saying Custer “screwed up” (Diagnostic Session). Rather, Terence said, the student thinks “the debate was: Was he real or was this some kind of imposter?” (Diagnostic Session). As she made her decision on the student’s fit to the class, Terence said, “I’ll always be on the lookout for poor little [1011-02-03] because she’s not a great reader, but I would definitely keep her in my class. I do not see any reason to bump her back anywhere. These are not significant problems” (Diagnostic Session). From what she said here, Terence’s plan was to watch after the student, rather than to push her down.

Again, noting concerns about students is not uncommon. Frequently teachers use the diagnostics not just to recheck placement, but to see students’ individual issues, and

as Smith (1992/2009) discussed, teachers may be reluctant to move even weak students.

In addition to these comments though, Terence also explained that she

[thinks the student] is having a problem with lens reading. She was asked to do too many things for this assignment. She had to read Graff and Birkenstein, apply it to this essay that was talking about two different scholars, and she just didn't understand it. (Diagnostic Session)

This explanation, that the student “just didn't understand” a too complex a task or reading, is a comment Terence discussed in her interview as well. And what she said there, in conjunction with this comment, shows for her that discerning who belongs in the course is more writing-focused than reading-focused.

When asked during the teacher interview what Terence felt about the placement prompt, she had a lot of comments on her dislike of the focus on sports in the Graff piece, the heavily excerpted nature of the passage, and the frustration expressed to her by students in her classes. She said, “I know students constantly every year, ‘Why did we have to talk about Pluto? That was so stupid! I didn't understand that at all’...and so I'm kind of thinking we should give them. If we want really accurate placement results, we need to be giving them something that they understand better. And I think that they don't understand this.” After reviewing more of the prompt and readings to refresh her memory, she later said, “So I always thought that this prompt was not fair. I don't think it's a fair way to assess their writing” (Terence Interview). Two things stick out, particularly in contrast to Odysseus's comments on both his prompt and the placement prompt. First, Terence believes the purpose of the placement is to determine writing ability alone, as she thinks that if students struggle to understand the reading or the

prompt, then the prompt would be an unfair assessment of their writing. And second, discovering which students misunderstand doesn't seem to be a determining variable of their placement, but interference to the determination.

The reason for her lack of worry about reading may be part of the kind of reading issues she is used to seeing. As I discussed extensively in the placement data, Course 11 students rarely misread overall. Instead they have moments of misreading that placement readers feel will be solved when they are forced to try to substantiate their reading with text. The idea is that they come face to face with the actual words of the text, and by being asked to pay attention to those words more closely, they fix their misreading.

In contrast, the biggest difficulty for placement readers—that is, when they are trying to decide placement on a difficult essay—comes down to determining if a student's reading problems will be solved easily through close work with the text, or if the student will continue to misread, struggle too much in the process of close reading, or require too much support. Making this determination for some students, those exposed as problematic in the placement analysis, is not easy, even for the placement readers, as there is often little on the pages of these students' essays to be sure. Thus, given Terence's course teaching experience with typical Course 11 students, students who rarely have misreading of a serious nature, she could believe that the misreading she notices in her students' diagnostic essays could be easily worked out as students were required to pay closer attention to the text.

And Terence makes a comment that suggests she may believe misreading can be worked out. On student 1011-02-02, Terence noted that the student was not “distinguishing between what Utley says and what Ambrose said” and that was “what we

were looking for in terms of the like the debate” (Diagnostic Session). In response though, she noted that she couldn’t recall any of her students who understood that, and she said,

when I think of [Course 11], I think of us as. That is what I am there to teach them. They don’t know how to do lens reading. They’re not necessarily already distinguishing between scholars who say different things. That for me, and I hope I’m not wrong about this, that would never be a reason for me to bump them back.
(Terence, Diagnostic Session)

If the course is going to show students how to read to understand tasks and differentiate voices in a text, then for Terence, noting that the students do not do these things is not concerning. This is an interesting contrast to Theresa and Odysseus, and to Course 11 placement, where there is a weighing of students’ reading issues against other factors to determine what the absence of these aspects might mean. There is a suggestion in the other Course 11 teachers’ readings of student work that reading issues could indicate a student will struggle in the course, but Terence makes no suggestion of that here.

Terence does not link the criteria of work with the text to the criteria of reading.

As Terence discusses students’ work with text, we see that she believes that the course will teach students these skills. Yet, we also see that Terence defines students’ work with the text in terms of knowing to use quotes, choosing and using them purposefully, and citing them properly, and this criterion is never directly linked by Terence to students’ strengths or difficulties with reading. For five of the six students she discussed on this criterion, Terence’s comments were positive and focused on how students had a quote or quotes: “Oh, he has a little quote here!” (1011-02-01, Diagnostic Session). For two of

these students, Terence commented, essentially saying that the student “knows the importance of quoting” (1011-02-03, Diagnostic Session), which suggested that for these students and for those noted simply as having a quote, Terence was pleased that they knew at least that much about working with text. Additionally though, working with the text for Terence is about understanding academic forms of citation and attribution. For example, with student 1011-02-06, Terence said, “Oh, so that’s a nice quote. Very good. And she almost documented it correctly. Almost” (Diagnostic Session).

In addition to knowing to use a quote at all and knowing how to cite a quote, working with the text means a writer purposefully chooses quotations. For example, as Terence continued to talk about student 1011-02-06, she added, “Okay good. So she has chosen. You know, two excellent quotes that sort of set these two positions against each other. Nice,” and she followed up this observation with “She knows how to use a quote” (Diagnostic Session). The student was essentially being praised here for the purpose behind her use—for the choice of these quotes in relation to what she was saying. Here, Terence approached an observation that could be connected to the student’s reading, as being able to choose helpful quotes presupposes you have a reading, but reading as a factor was not directly mentioned by Terence.

There was one additional moment where Terence’s comment on working with the text moved toward reading, on the aspect of explaining quotes, but not much was said in this moment to detail Terence’s judgment on the quality of the student’s explanation. Terence said, “‘Put in your oar.’ She’s explaining that” and followed it up with “What she’s doing so far with Graff and Birkenstein is really nice” (1011-02-05, Diagnostic Session). Here Terence described the student’s explanation as “nice,” but the relationship

between reading and work with the text that we find in Theresa and Odysseus's think-alouds was not shown here. In Terence's think-alouds, there is no suggestion that a student can work on misreading through direct work with the text. Nor do we see her doubt a student's reading if he or she fails to work with the text.

Terence's view of what it means to work with the text discussed so far here is backed up by her interview responses as well. When she was asked to talk about what students learn in her Course 11, Terence said the following about working with text: "[Students] are going to know how to set up a quote, how to use a quote, how to evaluate a quote, and how to document and cite that quote by the end of the class" (Terence, Interview). Terence was not focused on what students came to consider from or understand from the texts through their work with quotations, as Theresa and Odysseus were. She was more worried that they knew to have a purpose in the choice and use of quotes and that they knew the proper forms and styles of integrating quotations.

Terence's concerns, while not reading focused, are a large part of what the course is meant to teach students. The course objectives and departmental learning outcomes claim students will learn to use what they've read or found through research in their writing properly. Yet, for those instructors who used to teach Course 11A, the quality of a student's understanding is not separable from his work with the text, as is the case for Terence here.

Essay structure and grammar are bigger indicators of concern than other criteria for Terence. As I've said earlier, while Terence discussed and noted issues with reading, these rarely concerned her or were not mentioned as she made her determination of a student's fit to her course. Instead, she more frequently mentioned essay knowledge

or grammar when stating why she felt a student fit her course. For example, as Terence made the final decision on student 1011-02-01, she said, “Oh, yeah. This kid is definitely ready for [Course 11]. Definitely. He has ESL issues still, but it looks to me like he’s even overcoming most of those” (Diagnostic Session). Or with student 1011-02-04, Terence said, “She’s just not a good reader...She tries. And in [Course 11], if you have decent sentence structure, which she does. She can put together a paragraph. She knows the importance of talking about the scholars...And she has named them...she’s fine” (Diagnostic Session). Five of Terence’s eight students had their fit-to-course determination made as Terence noted some existing knowledge of essay structure, strong grammar and style skills, or only minor errors. For Terence, these are signals of students with prerequisite knowledge for the course.

Janet’s criteria. As I said above, not all Course 11 teachers have experience teaching the old Course 11A course. Janet, like Terence, has far more experience with Course 11 than she does Course 11A. As such, her think-aloud protocols show interesting differences in criteria and in the relationship of criteria in comparison to old Course 11A teachers, and they show similarities to Terence’s criteria. Table 15 displays the total number of students discussed by Janet for each criterion.

For Janet, the most frequently commented on criteria was the assignment task followed closely by a discussion of students’ framing and moves, which, as we will see, are two criteria she relates closely to each other. Further, while the criteria of reading and working with the text were discussed, we will see that for Janet, there is a difference in the relationship between these criteria than the relationship seen at placement.

Table 15

Total Number of Course 11 Students Discussed by Janet for Each Course Criterion

# of Students	% of Students	Name of Criterion
16/20	80%	Assignment task
14/20	70%	Framing/moves
12/20	60%	Reading issues/quality
12/20	60%	Working with text
12/20	60%	Length/amount of writing
9/20	45%	Essay/ paragraphing knowledge
9/20	45%	Grammar
6/20	30%	Language
6/20	30%	Thesis/project
6/20	30%	Use of required texts
6/20	30%	Thinking
4/20	20%	Print conventions
2/20	10%	Academic rhetorical conventions
0/20	0%	Writing process

Following the assignment is about making the right moves for Janet. Janet

frequently commented on the way students addressed the assignment. For example, with student 1011-09-01, Janet said,

instead of understanding that Graff and Birkenstein are a lens, he's gone into just discussing Graff and Birkenstein. And he doesn't understand that he's really to talk about, to write about Stephen Ambrose. He got bogged down in Graff and Birkenstein. He did talk about Stephen Ambrose though. (Diagnostic Session)

As we see here, Janet discussed the student's ability to address the assignment task, but she explained this in terms of the student making the right moves in the right amounts. Noticing the kind and amount of moves is something that comes up with other students. Another example was student 1011-05-01, where Janet said the student "didn't understand the lens at all. Just a little tip of the hat to the *They Say, I Say*" (Diagnostic Session), indicating the student didn't spend enough time on the first text to use it as a

lens. In fact, Janet made a comment on a students' sequence of moves in relation to the assignment on 13 of 16 students she discussed for assignment issues.

Assignment problems are not reading problems for Janet. The issue of students not following the assignment is typically seen differently by Janet than it is in the placement sessions or in the diagnostic readings of Theresa or Odysseus. When Janet notes that students have not followed the assignment, she attributes it to their not having looked carefully at her instructions, or their not understanding the assignment, presumably in terms of the moves they must make or the amounts they must write about for each text. For example, with student 1011-05-04, Janet said, "They are talking totally about Graff and Birkenstein. Again, I think this just screams... 'I know what this is all about. What am I doing in this course?' Because you didn't read it! It's clear that you didn't" (Diagnostic Session). As we see here, she attributed the student's issue with doing the assignment to thinking he knows what to do over reading the prompt. Janet did not consider that the student's lack of Ambrose could be a reading difficulty, where he struggled to read Ambrose or to read him in connection to Graff and Birkenstein. On the 16 students commented on for assignment issues, Janet discussed 10 of them as not reading or understanding the assignment, and on all 10 no connection was made between students' failure to do the assignment and any troubles they might have with reading.

This understanding of assignment issues may be explained given Janet's typical population of students. Course 11 students, as we saw in the unproblematic Course 11 placement discussion, usually attempt the task with some degree of success, and the issues they have at achieving success on the assignment are of a relatively minor order, perhaps they've not fully connected the two texts in depth or perhaps they only just

sketched a connection. Janet's sense that these students must not have paid attention as closely as they should to the assignment could be related to her experiences with students who do not typically fail to attempt the task.

Take for instance what she says in regards to student 1011-09-06. Janet said, "And this person is talking about Donald Trump. Not really sure why. And really not understanding the task at all" (Diagnostic Session). This student, far more than any other in her set, exhibited problematic behaviors in regards to assignment and reading, but as she says here, Janet was not quite sure what to make of what he's written. After reading a bit more, Janet said the student was "not putting much effort or much thinking into it," and, as we will see shortly, effort is the typical explanation for Janet.

Fixing misreading is about effort, length, and an investment of time. For Janet, reading is discussed, and she noted when students have misunderstood. For example, with student 1011-09-05, Janet said first that the student was just "reexplaining Graff and Birkenstein and not really understanding it ever." But then she read the student's claim out loud that "Custer proves that jogging out of your comfort zone will lead you to your future dreams," and she said this idea "is a little confusing because Custer ended up completely dead and all of the people with him. I hope that wasn't his future dream" (Diagnostic Session).

While reading issues are something Janet notes, the implication she makes to explain these issues often relate to the student taking more time. For instance, with the student above, Janet went on to say, "I don't think he read the whole thing. Obviously he didn't read enough to understand that Custer dragged a bunch of young men into the Battle of Little Big Horn and every single one of them died" (1011-09-05, Diagnostic

Session). Thus, the student's problem was he didn't read fully, and the course will teach him to fix this by pointing out that he has to invest the time to read the text completely, and once he does, he can figure out the problem.

There are other indications that Janet believes students can fix misreading with effort or the length of their responses. Student 1011-05-03, like his counterpart above, was described as "not really reading it completely," but the explanation Janet arrived at here is that the student was "not really wanting to put in that effort to actually read" (Diagnostic Session). And for Janet, student 1011-05-05 seems to need to say more to get to the right reading: "There is no understanding of at all of the argument that Ambrose is proposing. Just really, just giving a tip of the hat to Ambrose now that I've looked at it more deeply" (Diagnostic Session). Combining Janet's comments on the amount or length of space a student devotes to particular ideas with her sense that doing the assignment involves the student understanding the moves he or she must make, it seems Janet believes that the course will teach students to read more fully by asking them to write more content on the right moves, so they invest enough effort and time to figure out a text.

Proper placement means students with good sentences, coherent points (even if off topic) and an interest in thinking. While reading issues were frequently discussed by Janet, they were not as concerning to her, as students were seldom flagged as slightly below for these issues. Further, those students flagged as slightly below, 1011-09-06 and 011-09-09, were both flagged because they were not writing on the task, as we see with student 1011-09-09 when Janet said, "They are definitely low on the [Course 11] totem pole, so to speak, but they can be on it" after she noted that the piece was "off topic" with

no point of support, even for the topic they were discussing (Diagnostic Session). Thus the biggest indicator of weakness was being off the topic of assignment.

However, there seem to be other prerequisites for the course. As Janet discussed the majority of students (14 out of 20), she noted issues with meeting the assignment, having framing, or reading, but she then made a comment that suggests sentence structure, coherency, or interest were more key to why these students were still properly placed. Perhaps the best example is student 1011-09-11 who combined all three of Janet's course prerequisites. Janet noticed he was off topic, but the student said, as part of his essay, that he liked this kind of writing. In response, Janet said, "For a [Course 11], I am good with [being off topic] if they begin to think that there's something of interest. That is huge to me." She continued, noting as well that "He does have enough of a writing style. Enough of a coherence" (Diagnostic Session). Given the frequency of her comments on these criteria, I assume that if she saw problems here, she might be worried enough to push a student down. It seems unlikely that she would push anyone down for reading alone, and on this point she shares thinking with Terence. Given this divide between old Course 11A teachers and Course 11 ones—where reading can either be concerning enough to suggest a student may struggle or not really considered a key factor—one starts to wonder if these students do struggle or not?

Course 11 conclusions. Course 11 teachers have differing views of the importance of reading as a determining factor of placement into the course, and by extensions, they show differences in how they understand the work of teaching reading and writing to students in their Course 11 sections. Extending an observation drawn from an analysis of the placement sessions earlier, that problematic students are still primarily

suggested for placement with old Course 11A teachers, I believe the differences seen in the Course 11 teachers here might be related. These teachers were once completely divided in student experience, and still experience some degree of different student populations. Thus, it is possible that the divide in opinion on the importance of reading is related to these experience, or as Kane (2006) might say, the two sets of Course 11 teachers do not share *conceptual frameworks* on these students' behaviors.

Course 10 Teachers' Criteria

Now that I have described Course 11, I want to move to detail the criteria of Course 10. As I do so, I compare the course criteria of the individual teachers, and I compare these criteria to those discussed at placement. Specifically, I show there is a consistency in Course 10 teacher's criteria, particularly with reading and work with text issues, and I argue that this consistency has a lot to do with the stability of Course 10, despite the changes to the writing program. I also assert that Course 10 teacher's expertise can be potentially useful for helping the writing program understand the most concerning habits of reading, a necessary understanding if the program is to redistribute the old Course 11A student population into the two existing courses.

Adele's criteria. Since Adele is the Course 10 placement reader in this study, I start with her criteria. First, let me present, in Table 16, the total count of criteria discussed for the 10 participating students in her two sections of Course 10. Thinking back to Adele's Course 10 criteria, this frequency count is not all that dissimilar. During placement, Adele talked about the same four criteria that top this list, albeit in a slightly different order with the criterion of work with text falling in the fourth spot. Given that this set is based on 10 students while the placement reading considers 108, the

differences are likely a byproduct of the size of the group. Regardless, Adele's concerns, as we will see, are the same here as they were at placement. She was concerned at placement with the students' reading, and she was concerned about their understanding of the assignment, as well as their work with the text, and both of these concerns were considered related to reading.

Table 16

Total Number of Course 10 Students Discussed by Adele for Each Course Criterion

# of Students	% of Students	Name of Criterion
10/10	100%	Reading issues/quality
9/10	90%	Working with text
8/10	80%	Assignment task
8/10	80%	Essay/ paragraphing knowledge
5/10	50%	Framing/moves
3/10	30%	Grammar
3/10	30%	Academic rhetorical conventions
2/10	20%	Language
1/10	10%	Length/amount of writing
1/10	10%	Use of required texts
1/10	10%	Thinking
0/10	0%	Thesis/project
0/10	0%	Print conventions
0/10	0%	Writing process

Reading is the key criteria for Adele, and it relates to a student's work with text.

All of Adele's students were assessed in terms of reading. While other criteria were discussed for each student, reading was Adele's primary focus. In fact, while Adele would mention criteria like paragraphing or grammar at the beginning or end, or even intermittently as she noticed issues, the majority of her discussion throughout, as she read and made sense of each student's piece, was on their reading and how they worked with text.

For Adele, a student's difficulty reading the text was related to his or her work with the text, and together these issues could suggest a student was unprepared for reading at the academic level. With student 1010-07-05, Adele noted precisely the student's reading level when she realized the student had written primarily an informational retelling of the facts provided about the texts on the prompt rather than doing any direct work with pieces of the text: "She like focuses on content, and my guess is that's what she's been taught to do. It's that level of reading" (Diagnostic Session). Continuing, Adele described the relationship of one's reading to the way one works with text in her characteristic phrasing of "She just doesn't know what to do with what she read" (Diagnostic Session). To come to a reading that was more than just the "content" level, the student must do something with the text, and Adele suggested this student does not yet know how to "do" something, the implication being the student must do something with the ideas in her mind or on the page to come to a reading, and because she doesn't, she only finds information to recount in texts.

For Adele, how a student reads is key to seeing if they will struggle to do academic work, and how a student works with text is the way a student gets to a level of reading past the content one. If a student cannot work with the text, as 1010-07-05 cannot, then when they do offer some thought in response to the text, it will be one that has not really heard and participated with the text. It will be one that is "off" or "cliché," as Adele labels such responses, such as when 1010-07-05 insists the texts together show us to "stand up for what we believe" just like General Custer did.

Adele considers the assignment a reading task. When asked in her interview to talk about the placement prompt, what it asked students to do, and how it related to the

diagnostic prompt she uses, Adele made it clear that both assignments ask students to do the same work, and that that work involves reading in a new way. She said, “What the [placement] prompt is asking them to do” is

use those two texts because it is not a standard compare these two texts in terms of how they are the same or different. I think the most difficult thing for the students is that it asks them to use one text to read the other. Like to apply this one reading, Graff’s argument, to the second, Pluto article. That, to me, is the key thing here. (Adele Interview)

Adele emphasized the way the prompt asked students to read, or at least to use their reading of one text in relation to the other in a specific way.

As she continued, Adele described where students in Course 10 will struggle with this task. She said, “First of all, they are going to struggle with this ‘Hidden Intellectualism’ text. Okay, because it is academic. So they are going to link it to something that they can connect with. Sports” (Adele Interview). The point here for Adele was that the first text was likely already difficult for the students without their also being asked to read the other text through it. Then she described what, for most, is the key problem—the problem of reading through another text. She said, “Then, the biggest thing they are going to have difficulty with is figuring out how to use Graff’s text to read what is happening in this Pluto article” (Adele Interview).

Since she knows Course 10 students will struggle, she also described possible strategies Course 10 students resort to when they are not sure how to read at these academic levels and in these academic ways. She said, “So for some of them, they will just summarize Pluto because they can understand that” (Adele Interview). Others, she

said, who struggle to read one text through the other, will summarize “both” texts and then “talk about do they agree or not that Pluto should be a planet” (Adele Interview). Adele’s interview makes it clear that some of the comments she makes when she discusses students’ responses are about reading even as she describes what seems like aspects of writing—the moves the student makes or their meeting the assignment.

For example, with student 1010-07-03, Adele discovered after reading a few lines that “This is a summary,” and she went through the subsequent paragraphs quickly, saying “Custer. Custer. Custer. All about Custer” (Diagnostic Session). Eventually the student said something about critical thinking, and Adele described the student as “now desperately trying to pull in the other passage” (Diagnostic Session). On the surface, these comments seem about the moves of the students’ writing and their ability to meet the assignment, but given Adele’s interview response, we see she is assessing their reading in these moments as well because the assignment for her presents a reading task. A student who resorts to summary, for Adele, is one who is struggling to read in academic ways.

Comments like these are not uncommon. Again, with 1010-07-08, Adele noted the student has approached the task as a compare and contrast: “Here’s how they’re similar and here’s how they’re different” (Diagnostic Session). Given Adele’s interview, we know she thinks of this problem as, at least in part, a reading problem. Given this view, it is clear to see that even as she discusses a student’s ability to do the assignment, Adele is talking about their reading level.

Adele also considers essay and paragraphing. One of the first observations Adele makes on eight out of 10 students has to do with the students’ paragraphing, and

she seems to scan the responses early on to see what the students know about paragraphing and essay structure: “Paragraphs. So that tells me something. Knows how to structure an essay somewhat” (1010-02-03, Diagnostic Session). For Adele, a student who has essay knowledge, particularly moving past just paragraphing—having an introduction or conclusion for instance—is a more prepared student academically although essay knowledge is not considered in isolation from their reading. This is best shown by comparing Adele’s comments on two different students who both present a high level of essay knowledge.

Student 1010-07-02 had “five paragraphs! Okay. So she’s got this introduction” (Diagnostic Session). And as Adele read, she noticed the student tries to work with text, using a quotation, but largely misreads even the easier text by Graff and Birkenstein. When deciding where she would rank this student on the survey of adequacy, determining her fit to the course, she said, “I would put her. She has like some good writing skills. She has understanding of the conventions. I would probably put her slightly above. I think she needs to work on the reading” (Diagnostic Session). The students’ preparation on essay knowledge and even the conventions of grammar, mean she is more prepared for college level work in ways, but her misreading of both texts mean she may struggle, at least enough to warrant her remaining in the course.

Student 1010-07-05 has a similar combination of strong essay knowledge, with “Five paragraphs” and “the introduction” (Diagnostic Session), and reading problems, but as Adele assessed this student’s fit to the course, she said, “I would say she is a prototype even though she has the paragraphing and things” (Diagnostic Session). The difference between these students seems to be that the majority of discussion on student 1010-07-05,

as we saw a little earlier, noted her focus on information and taking more of that information from the assignment prompt than even the text. In comparison, student 1010-07-02 tried to work more with the text directly and didn't stay at the level of content summary. In short, essay and paragraphing knowledge are important skills for college level work, but the designation of a student as a Course 10 student involves determining the student's reading skills and level.

Daria's criteria. Daria, who also has two sections of Course 10, shares many of the same deciding Course 10 criteria. Table 17 shows the total count of criteria discussed for participating students in Daria's two sections.

Table 17

Total Number of Course 10 Students Discussed by Daria for Each Course Criterion

# of Students	% of Students	Name of Criterion
21/25	84%	Reading issues/quality
18/25	72%	Working with text
14/25	56%	Essay/ paragraphing knowledge
13/25	52%	Grammar
12/25	48%	Assignment task
10/25	40%	Thinking
6/25	24%	Length/amount of writing
5/25	20%	Framing/moves
2/25	8%	Use of required texts
2/25	8%	Academic rhetorical conventions
1/25	4%	Language
1/25	4%	Thesis/project
1/25	4%	Print conventions
0/25	0%	Writing process

For Daria, the most frequently commented on criterion was reading issues followed closely by a discussion of students' work with text, which, as we will see, are two criteria Daria relates to each other. Like Adele, Daria still considers a student's essay and paragraphing knowledge, but that often doesn't change her concern about what she

finds on a student's reading. Unlike Adele, however, Daria noted more students with grammar issues, which is likely related to the larger number of ESL students participating from her sections in comparison to Adele's sections, since grammar was something Adele considered when she was a placement reader for Course 10 over the summer.

Daria sees reading as key and work with the text as a part of reading. The most frequently discussed issue on students' writing was again reading quality. Daria was concerned by a number of reading behaviors—in particular the habit of many students to turn to clichés, generalizations, and assumptions. For example, Daria saw three students, 1010-01-10, 1010-06-04, and 1010-06-12, as all coming to a misreading of the Ambrose text by making an assumption about what teacher's value and assuming that the text must represent an example of a good critical thinker. For example, about student 1010-01-10, Daria said, "He made an assumption that 'Hey, they value critical thinking. This other piece is talking about this guy. They value it. They're giving me this other piece. He must be a good critical thinker.' So he's not reading carefully" (Diagnostic Session).

Other students read to clichés, and what Daria said about some of these students suggests the relationship between reading and work with the text for her. About student 1010-01-04, Daria said, "This is a standard response. Be a go getter. Everybody should struggle to get what they want. For me, that makes her mainstream. Because she never really touches down on the text" (Diagnostic Session). This student makes the point of both texts into whatever clichéd advice she's heard, and Daria believes this was caused by her not working with the text directly. "Touch[ing] down on the text" would mean locating a moment inside the text to show what she means, and the student doesn't know

to do that, either in her thinking or on the page, so she comes to what she already believes about the world instead of a reading that represents the text.

The way Daria described student 1010-01-09's reading issues suggest he has more experiences with "moving from book to world" than from "text to text"—that is, he is familiar with relating ideas to his experiences, and these have helped him make meaning and address assignments in the past (Diagnostic Session). But in regards to academic work, these strategies take him away from reading closely. We see this as Daria described the student as writing a thesis that argues that you have to do certain things to become a critical thinker, and the list of things he gave were his own observations about reading newspapers and books. Daria said, "He had materials to go to. He could go to Custer. He could use [Graff and Birkenstein]. But he stays in this general space...He's not quite sure how to use the text to get said what he wants to get said" (Diagnostic Session). As Daria assesses the reading of her students, she looks for reading habits that suggest students will look away from the texts they encounter, and by extension, miss out on what the texts can add to their knowledge and thinking.

Daria notes essay and paragraphing knowledge. The majority of comments Daria made on essay and paragraphing knowledge seem to simply be her noting how many paragraphs students have. For 12 of the 14 students commented on for this criteria, the comment was simply something like "one long paragraph" (1010-06-04, Diagnostic Session). So on the surface, it seems this criterion is only an attribute she takes note of, but doesn't factor into her fit-to-course determination. But there were a few moments where the student's essay and paragraphing knowledge was discussed in relation to Daria's fit to course determination.

For student 1010-06-06, Daria saw early on that the student had written only a “single paragraph. No title,” which suggested lack of essay knowledge (Diagnostic Session). When she read the student’s response, she noted the student was “in the ballpark of” Graff and Birkenstein’s text, although “not using, so much, their language,” but she did quote “an important moment” from Ambrose that helped her not misread and think “Ambrose and Utley [who Ambrose cites] agree” when they don’t. So, the student does well at reading. Yet, when Daria ended her discussion of this student, turning her attention to deciding where to place the student on the survey of adequacy, she said, “She does a good job reading both pieces, but she responds with just one paragraph” and “has a high level of sentence-level error and sentence boundary issues” (Diagnostic Session). This student’s strong reading was considered against essay knowledge (and grammar). The same situation happens with student 1010-06-07 who has stronger reading, although not as strong as the first student, but can’t fit his response into accepted the “academic forms and formats” of essay structure (Diagnostic Session).

Daria also weighs grammar and language problems against other criteria. As we saw with student 1010-06-07 above, Daria weighs a consideration of a student’s grammar issues against the criteria of reading, work with text, and essay knowledge. And she does this with students whether they are native speakers, like the student above, or ESL students. But she did have a larger number of ESL students participating in the study than other teachers, so it was not surprising that the percentage of students who had comments made about their grammar issues was higher for Daria than it was for Adele.

Seven of Daria’s 25 participating students were ESL students, and what she said about the majority of these students showed that despite having ESL issues noted, they

frequently had similar issues to their American peers that place them in the Course—they struggle with reading and how to work with text. Occasionally, a student would be a little stronger with reading but present Daria with worry about their ability to handle academic work without support and additional time for their ESL grammar issues. This is true for student 1010-06-01 who Daria described as having quotes and “[recognizing] that there was a time when Custer was exulted by many. And then came a time when he was exulted only by a few,” but also “repeating” the text at times (Diagnostic Session). For Daria, the reading was strong, as the student was “struggling to read well” (a phrase that suggested the student was reading well although exerting effort to do so), but the student had “tons of ESL interference” (Diagnostic Session). What we see is that Daria often factors grammar into her determination for both native and non-native students, particularly when they are stronger readers than typically seen in the course.

Hannah’s criteria. Hannah discusses many of the same criteria as her fellow Course 10 instructors. Table 18 shows the total count of criteria Hannah discussed for participating students. Like her counterparts, Hannah most frequently comments on reading issues and work with text, as well as the assignment task. But there are some slight differences in what Hannah says in regards to these criteria, at least as gleaned from her diagnostics. There are a number of explanations for the differences. First, Hannah has been teaching in the program for the fewest number of years, with just three years at the time of this study (Hannah, Interview). Having never taught any other course in the program, and having never experienced the writing program with a Course 11A option, it is possible that Hannah is not impacted by the reading distinctions that used to

exist between the courses in the old course sequence. Thus she may not consider reading as significant to her fit to course determinations.

Table 18

Total Number of Course 10 Students Discussed by Hannah for Each Course Criterion

# of Students	% of Students	Name of Criterion
8/9	88.89%	Reading issues/quality
8/9	88.89%	Working with text
6/9	66.67%	Assignment task
6/9	66.67%	Grammar
6/9	66.67%	Framing/moves
4/9	44.44%	Essay/ paragraphing knowledge
2/9	22.22%	Language
2/9	22.22%	Academic rhetorical conventions
1/9	11.11%	Length/amount of writing
1/9	11.11%	Thesis/project
0/9	0%	Use of required texts
0/9	0%	Thinking
0/9	0%	Print conventions
0/9	0%	Writing process

Second, Hannah has a relatively smaller number of students' participating but with a higher percentage (44.44%) of those students being flagged as slightly above, so it was more difficult to see what she would say about a larger variety of prototypical students. Regardless of the slight differences though, Hannah's diagnostic combined with her interview responses show a large amount of similarity to her fellow Course 10 colleagues.

Hannah notes students' reading, but it may not be her main concern. Hannah often simply notes misreading or odd reading while reading students' diagnostics, but she doesn't speculate on or categorize these reading issues as much as other Course 10 teachers. For example, on student 1010-05-06, Hannah indicated just that the student had some "slippage" on "the purpose of Ambrose's essay," not saying much else on what that

meant exactly or what it showed about the student's preparation for the work of the course. On student 1010-05-02, Hannah said the student "understand[s] what Ambrose's piece is about" but "is spending much more time focusing on Custer" than what it all means for "critical thinking" (Diagnostic Session). Where one of her Course 10 colleagues would likely have gone on to think about what that could show about the student's ability to read at the level of connecting the two texts, Hannah just moved forward. She did at times note one particular subcategory or issue with reading. When she discussed three separate students in the bunch, of which 1010-05-04 is a good example, Hannah described an issue with "confusion about who wrote what" (1010-05-04, Diagnostic Session), a reading aspect she also noted as something she looked for in reading diagnostics when interviewed. Given the number of names inside both Graff and Birkenstein's text and Ambrose's text, Hannah points out when students show some issues keeping track, although it doesn't usually seem too concerning to her, as 1010-05-04 was still ranked at slightly above, even with this issue.

Despite the lack of reading issue categorization, there are some indications in Hannah's interview response that reading at the level of achieving a strong connection between the two texts is important to her sense that a student is more or less prepared for the course, although it is still not as worrisome as it would be to her Course 10 colleagues. When comparing the placement prompt, which she had never seen, to the diagnostic one she used from the department, she stressed that both test "[students] ability to take two seemingly disparate texts and bring them together in some kind of creative, intelligent, cohesive way" (Hannah Interview). Despite the emphasis on connecting, Hannah didn't think the responses should be assessed with one answer in mind, with

“right or wrong” readings, because it was possible for students to “misread it, but in very interesting ways” where some readings would be “very promising misreading[s] at some level” (Hannah Interview). These comments by Hannah seem to stress the importance of remaining open to multiple interpretations on the connection between the texts.

But there are also indications that there are better readings for Hannah, as when she was asked where her students would struggle with the task of putting the two texts together, she replied that “When we did the diagnostic, you saw that several of my students were kind of on the high end, and some...were on the low end. The ones on the high end, they kind of get the relationship. The ones on the low end have no idea how to bring these two texts together. They just don’t see anything happening at all” (Hannah Interview). The wording here of “the relationship” in regards to higher end students indicates that what separates higher or lower students does have something to do with how fair their connection is to the intended meaning of these two texts and the prompt question.

A student’s work with text appears to be the main concern. As I’ve indicated above, reading issues, at least as represented in the participating diagnostics, do not seem to be what decides students’ fit to the course rankings for Hannah. However, what Hannah says about students’ work with the text does appear to be her main concern. Eight of nine students were discussed for their work with text, and Hannah considered such things as how well the student formed paraphrase and summary (1010-05-05), if they used quotes and how much (1010-05-03) or little they explained them (1010-05-07), if they integrated quotes well (1010-05-06), or if they did something with the language of quotes, aka “running with the language” (1010-05-04).

In addition to how frequently she discussed this criterion, and how much of her think-aloud talk was devoted to it, she also saw work with text as something that students must come to understand and its presence as an indication that the student was slightly above the course. The best example of this was student 1010-05-06, who was described as “actively quoting the text, working with it,” and also “doing so fairly deftly” with “lots and lots” of moments of quotation that have a “focus” (Diagnostic Session).

Work with the text will lead students to be stronger readers and to do stronger work of their own. On work with text, Hannah definitely links what students do with the text to their reading of it like other Course 10 teachers. On student 1010-05-07, Hannah said the student had “one good quote” that he used well with Ambrose, but then “[started] slippery-sliding around” after that, suggesting that the student was “almost working with the text” but still needed more pushing (Diagnostic Session). And student 1010-05-01 was “tangled up in the number of proper names and references” so much that she never really made a connection between all that she said about the Civil War and the critical thinking idea from Graff and Birkenstein’s text (Diagnostic Session). Hannah described this as the student not “engaging with the text” as she used it (Diagnostic Session).

The link between reading and work with the text is seconded by Hannah’s interview responses. In the interview, she described the course as teaching students to “work actively with the text” and to “do some interesting work with the text. Not just summarizing, but really sort of thinking about what the text implies” (Hannah Interview). She also said this work would require students to “pay attention to details” and “grapple with [texts] and grasp [them]” (Hannah Interview). Here her emphasis was that students

would learn to read and to do something of their own with texts by looking at the details and spending time trying to figure out what they mean.

Theresa's criteria. Theresa is the final Course 10 instructor, and interestingly, the only teacher who was teaching both Course 10 and Course 11 the semester of study. She was also the Course 11 placement reader. Table 19 gives the total count of criteria discussed for participating students in Theresa's section. The most frequently commented on criteria for Theresa with Course 10 students was issues with the assignment task followed closely by a discussion of students' reading and work with text. While the change to the criteria in this count total appears, on the surface, to suggest differences between Theresa and her fellow Course 10 teachers, we see that Theresa links the assignment to reading during the Course just as she did in placement. Thus, these comments are most often a form of assessing the students' reading for Theresa.

Table 19

Total Number of Course 10 Students Discussed by Theresa for Each Course Criterion

# of Students	% of Students	Name of Criterion
10/11	90.91%	Assignment task
9/11	81.82%	Reading issues/quality
8/11	72.73%	Working with text
5/11	45.45%	Use of required texts
4/11	36.36%	Academic rhetorical conventions
3/11	27.27%	Essay/ paragraphing knowledge
3/11	27.27%	Grammar
3/11	27.27%	Length/amount of writing
2/11	18.18%	Framing/moves
1/11	9.09%	Thinking
1/11	9.09%	Print conventions
1/11	9.09%	Thesis/project
0/11	0%	Language
0/11	0%	Writing process

Theresa sees the assignment as a reading task. When asked to talk about the placement prompt and its relation to the diagnostic during the interview, Theresa made it clear that the assignment she had written—for both placement and the diagnostic—was meant to elicit information about reading. Describing the placement prompt, she said, “It’s a text on text...We’re trying to see in this placement what happens when you give students a lens to look through.” Elaborating and discussing why she chose the passages she did, she said, “My thinking when I chose [a journalistic piece] was you’ve got the more complex text [Graff] already. Let’s see what they can do with that” in comparison to the easier text. She described a series of problems she sees typically saying,

There’s so many ways you can see what kids can do. Can they read one text through another? Are they critical readers?...Are they able to read and interpret a more complex passage as well as a more straightforward journalistic passage...I looked for passages that had quotations in them to see if they were able to manage text boundaries inside a text. If they understood when someone was quoting something and when someone was speaking...as the author from an authorial voice. (Theresa Interview)

From the choice of task to the choice of texts, Theresa had designed the prompt to show her what students could do as readers and where they would fall down.

Yet we also see this concern with reading in how she talked about student’s performance of the assignment. For example, with student 1010-04-06, Theresa noted that the student had written separate answers to what he had conceived of as separate questions from the prompt. She said, “Then he goes on to answer the third question, what does it mean to be a critical thinker. So he’s a literal kid...The whole notion of reading is

finding a main idea or answering a question,” that is “his job” (Diagnostic Session). She stresses here that the way the student thinks about reading is what keeps him from doing the assignment. He can’t connect the texts if he imagines the work of reading is to address questions separately.

Student’s conceptions of reading, in fact, are what Theresa usually discusses as she talks about their responses being off the assignment. About student 1010-04-10, the conception was that texts are information. Theresa said,

So she’s actually read Ambrose and kind of got the point. But she doesn’t understand Ambrose as a writer. Or Utley as someone who is being quoted or what the debate is over Custer. And so when she comes down to actually talking about the critical intellectual thinker, she makes Custer the great intellectual thinker. (Diagnostic Session)

For Theresa then, discussion about the assignment is really discussion about reading at an abstract or idea level and making connections.

Theresa’s concern is primarily with reading. Even when not discussing the abstract or idea reading level—that is the level of reading that Theresa relates to the assignment task—Theresa is still assessing student’s reading and seeing what their reading suggests about their potential for struggle. In the process of reading her students’ diagnostics, she categorized several habits of reading that typically cause students problems. Many students begin their writing by “making things up” (1010-04-05) from their own head about a topic, despite being given readings, or they change the topic of the text to “something [they’re] familiar with” (1010-04-13) so they can more easily write a response. Further, students may continue to make things up even when they claim to be

discussing the content of the text, what Theresa called “approximating” (1010-04-06).

These habits of reading were important indicators of not only reading problems that might interfere with students’ transition to college-level work, but also, as we will see in the next section, tell-tale signs of their unfamiliarity with working with the text directly.

Student’s come to a reading and show their reading in how they work with the text. For a student to show their reading, they have to be able to effectively use text. Take for instance, student 1010-04-12, who Theresa described as “[getting] the big idea about Graff” but not being able to “manipulate the text at all” (Diagnostic Session). Theresa believed the student had some sense, but they weren’t used to the academic work of displaying their understanding for a reader. While this occurrence was common in Theresa’s diagnostics, the more frequent issue in regards to Course 10 students is how their lack of work with the text or their ineffective work with it actually keeps them from coming to a reading. So a student, like 1010-04-11, who Theresa described as thinking of reading as content, “eliminates almost everything in the [text] that isn’t content about Custer” and ends up with “a story of why did Custer fail” (Diagnostic Session). Or a student can get stuck “repeating” at best (1010-04-06), or “copying” at worst (1010-04-09), and end up with no reading at all. Occasionally the opposite situation is true, where a moment of strong textual work shows a “sensitive thinker” lost somewhere inside a response that is largely “making [Graff] up” and “avoiding Ambrose,” as is the case for 1010-04-08 (Diagnostic Session).

Course 10 conclusions. Looking at all the Course 10 teachers, we see a largely shared consensus on key criteria, with the emphasis being primarily on either reading or some other criteria that the reader closely links to reading. Regardless of which of a few

criteria seems to make the determination, the things teachers say about them indicate that these teachers share a vision for the work of the course, and that the vision they share is also displayed through the placement criteria. All the teachers indicate that Course 10 students will be taught to work with the text more directly and closely, as that is not their habit and as they have little practice doing it. And all indicate that such work will help students learn to build a reading of texts that are difficult. When they find students who already have stronger reading and some skill at working with text, Course 10 teachers are typically concerned enough about students' lack of essay form or their frequent grammar errors to make them think the student would need additional instructional time to learn these lessons.

In addition to a large consensus on the importance of reading or reading-related criteria, all four Course 10 teachers make some observations, and three out of the four teachers make multiple observations on specific reading problems—issues with clichés and assumptions, a focus on content over idea, trouble tracking references to multiple voices in the text, and making things up that weren't in the text. These same distinctions happen at placement and are useful for the Course 10 reader to distinguish between students who have more concerning reading problems and those who have less concerning reading problems. In short, the expertise of Course 10 teachers includes a strong sense of what can go wrong with a student's reading.

Linking Course Conclusions Back to Placement

Having analyzed the course criteria, I have displayed that Course 11 teachers are split in the emphasis and weight they place on a students' reading ability in determining a student's fit to the course. In addition, I have suggested that the differences in criteria

may be related to differences in teacher's backgrounds as former Course 11A teachers or not. Alternatively, Course 10 teachers are much more united on the importance they place on reading as a criterion. They may also consider other criteria that differ from their colleagues, but they consistently relate reading to their other analyzed criteria or factor it in quite heavily in determining their concern for students in the program.

Where applicable I have made connections back to placement criteria, noting that Course 11 teachers, because of the split, are not all aligned with the key placement criteria. While all the Course 11 teachers see that one of the goals of Course 11 is to teach students about close reading, only half of them assess or make differentiations in their assessments of a students' reading when determining if a student is ready for their course. In contrast, Course 10 teachers share the criteria of placement and all show some noting of reading subcategories in their depictions of students' reading. These subcategories or differentiations are similar to those detailed at placement.

While I have made these connections between course and placement criteria, before going forward, I want to draw a connection between the findings of each of these sections as well. During my analysis of the placement criteria, I uncovered that Cardinal College could not really claim they were following the process of the expert-reader model (Smith 1992/2009, 1993) since the Course 11 reader often felt pulled between the reality of the loss of Course 11A, but her concern for the student population it used to serve. As a result, she would determine a student was Course 11A, and then she would weigh evidence for either Course 11 or Course 10 placement rather than immediately push concerning Course 11A students down to Course 10 when she determined they were not fully prepared. And, she would often suggest an old Course 11A teacher for any

student she decided to place into Course 11 who was problematic. This change in process cannot really be said to be fully Smith's expert reader model, yet I believe the alteration was made in the spirit of what Smith's process is built on—a concern with making the most adequate decision for a student's placement, and a desire to tap into what teacher's know about students in real contexts of learning.

Yet even Theresa admits that the result was not a structural solution, and as we have seen, weighing what was in essence already a fairly mixed set of characteristics produced inconsistent decisions with 18 different decisions made on students between the two reading sessions. Theresa's worry about pushing more students down is in keeping with her concern for adequate decisions, in the Smith (1992/2009, 1993) sense, and in the sense of Messick (1989a). These kinds of students were once able to be successful college readers and writers with four credits, and the burden of six credits instead—not to mention the stigma and disruption to their first year schedules and progress—is a real concern. Before moving to consider what all the data sets suggest about problematic placement students, I want to return to placement and use an example from the think-aloud records to plant a seed about the value of having conversation on the students who are most problematic, particularly when those conversations allow the two expert readers to develop and present a set of evidence for their concerns and put them forth to each other, rather than carrying that burden alone. Such conversations could tap into the stability of Course 10 understandings and the program could use those understandings to determine which student-concerns from the old Course 11A might present the most problems, resetting the course boundaries and developing a structural solution to the problem.

Student 1010-05-03 was a tweener in Redo Session 3. Upon her initial reading, Theresa's concern with the student was that she does an "approximation instead of close reading" so ends up with a "highly reductive reading" of Graff as "current events." She feels the student is on the border of Course 11, but decides not to place her in. Adele, in her reading of the essay, was also concerned. She felt the student had essay problems and wasn't a fabulous reader, but she was "actually trying to do a little bit of work with the text" and even to "say something," losing it as she moved to Pluto. She decided to push the student up. When both readers came together to discuss tweeners, they had a conversation about their differing sense of this student's reading, Theresa describing the reading as a reduction to "current events," and Adele making a case for another interpretation of what the student was saying.

The student wrote "it's important to take part in the happenings of the world," and Adele claimed that while this sounded like a reduction of Graff, it could also be that "sometimes with these students, they use words that don't really reflect what they're meaning because they don't have other words to use." This interpretation of the student struck a chord with Theresa, and she responded, "That's right. Exactly." Still, Theresa felt that while she could put the student into Course 11, as "She can probably do it," she also believed "She's gonna have some problems because for her, reading is too much closing the book and then remembering things. You know, her impressions" (Redo Session 3).

This dialogue is an interesting example of the way that conversation on a problematic student's essay can lend an opportunity for both readers to consider the evidence on the other side without having to, in one moment, produce both readings

themselves. It can also present a way for old Course 11A knowledge of our student population not to be lost, but to be folded into the understandings of the existing courses through time and further conversations.

Findings on the Link Between Placement Criteria and Course Criteria

Course 11 teachers' criteria differ significantly from each other, and for some teachers, this also means they differ from the placement criteria. It seems that if a Course 11 teacher's past experiences with teaching were primarily with Course 11 students, their assessment of how well students fit Course 11 shows less concern for reading. But, if a Course 11 teacher has had more experiences teaching Course 11A students, when the course existed, their assessment of how well students fit Course 11 shows more concern for reading. Alternatively, there is a consistency in Course 10 teacher's criteria, particularly on assessing students' reading and work with text issues, and this consistency may have to do with the stability of Course 10.

The take away from these differences has a lot to do with how teaching expertise is built and used as I have described it in Chapter Two. To build expertise, I argued, teachers must see a number of students in a course to gather a sense of what is typical in the given context of a course (Stock & Robinson, 1987). In this case, teachers are just like all readers: they must actively construct meaning while reading and they have to have practice with different texts to learn to adjust their interpretations for new scenarios of reading (Tierney & Pearson, 1983). Furthermore, once this sense of what is typical is established, it is likely that a large number of students must be seen to challenge that view and signal to a teacher that something else must be accounted for in their thinking, in their *conceptual framework* (Kane, 2006). If a teacher rarely sees certain issues or

primarily has certain experiences with students, these issues will appear idiosyncratic rather than central, and they will cause no change to the framework from which the teacher operates.

When a teacher plans the work of a course or assesses work in a course, Kane (2006) said teachers are drawing on their *conceptual framework*. And since Cardinal College, building on Smith (1992/2009, 1993), seeks to tap into teacher's conceptions to make placement decisions, when teachers assess placements or diagnostics, the goal is also to have them draw on the expertise of their conceptual frameworks. Given the stability in the shared conceptual framework of Course 10 teachers, Course 10 teacher's expertise can be potentially useful for helping the writing program understand the most concerning habits of reading, a necessary understanding if the program is to redistribute the old Course 11A student population into the two existing courses.

It is important to note that Kane asserted that teacher's *conceptual frameworks* would need to be evaluated and he included, in a list of what that takes, a consideration of the teacher's evidence, their credentials, and their experience. In the discussion so far, my focus has been on past experiences that these teachers do not share. As I go forward, I evaluate the conceptual frameworks represented by teachers further, looking at the data on particular student groups across that data sets to see, among other observations, what that evidence suggests about if and how students flagged with reading issues at placement do go on to have difficulty fitting into these courses.

Are Placement Concerns and Decisions Linked to Course Concerns and Decisions?

To consider the last subquestion of the study regarding the link between expert readers' placement criteria and course decisions, I looked at the records for those students

who were flagged by readers as falling in between the courses (old Course 11A students), those whose redo session decision was different from their original placement decision, and those students who were session tweeners. Using information gained from all think-aloud protocols, the adequacy surveys, and the end of course results sheets, I tracked these edge students across both the placement and the course sides of the data, checking for indications of possible links between course designations and outcomes and placement ones. There are a number of important observations that came out of this analysis; however, it is important to say first that this study did not collect the right data to know with certainty what exact issues factored into each student's course outcome. While teachers submitted a final note on each student, for many teachers these notes were brief, and as such were likely not a full view or accounting of student performances. Given this fact, some of my observations, particularly on what can influence students' final course grades and outcomes, are speculative, requiring future study to confirm.

This section primarily discusses how the relevancy of placement designations and concerns to teacher's course designations and concerns is different for problematic Course 10 and Course 11 students. First, for problem students in Course 10, there is a clear relationship between problematic placement designations and discussion and the teacher's course concerns and discussion. Course 10 students with problematic placement designations indicating one or both readers felt the student may be above Course 10 were frequently either ranked as slightly high for the course overall by the course teacher or were ranked as slightly high on the same criteria that the placement reader noted. Additionally, while a large majority of Course 10 students overall did well in the course, these problematic Course 10 students, taken as a whole, outperformed their peers in terms

of final grades. Finally, while the majority of Course 10 students did well in the course, I will suggest that Cardinal College has to consider that Course 10 could be demotivating for some students, as a few students designated as being above the course did not do as would have been predicted given their placement and diagnostic records.

Second, for Course 11 students, the criteria discussed at placement are not as concerning to course teachers. Only once was a participating Course 11 student with a concerning placement designation also ranked as slightly low for the course overall by the teacher. However, teachers frequently noted the same issues on students' diagnostics that were noted as concerns during placement, but Course 11 teachers rarely saw these issues as serious concerns. Despite the difference between the placement reader's view and how Course 11 teachers saw these criteria during diagnostic review, problematic Course 11 students taken as a whole went on to underperform in relation to their unproblematic peers in terms of final grades. The difference in performance between problematic Course 11 and unproblematic Course 11 students and its alignment with placement designations suggest that while the placement concerns for problematic Course 11 students are not shared by Course 11 teachers, something being noted at placement is relevant to the course outcome. Further, I suggest that even though the majority of these problematic Course 11 students are able to pass the course regardless of how they compare to their peers, the comparatively low grades should be concerning, as students who fall behind in Course 11 face more difficult tasks and less support in Course 12 going forward.

Problematic Course 10

Course teachers are typically reluctant to move students from their courses (Smith 1992/2009), and thus when we factor this in, it is logical that a teacher's level of expressed concern will be of a lower degree than that expressed by placement readers. Course 10 teachers, while not suggesting that the 30 problematic Course 10 students were improperly placed, do flag 15 of these problematic students in some way and make a final course decision on another three that suggests a connection between placement and the course.

Eight of the problematic Course 10 students were flagged as slightly high for the course (with three of these students going on to jump Course 11 and go straight to Course 12). For example, student 1010-04-11 was flagged as Course 11A by both Theresa and Adele at placement, but was put into Course 10 during both placement sessions. However, during the redo placement session, the student was tweened. As the course teacher, Theresa, read the diagnostic essay, she noted at first that this student showed signs of being a Course 10 or a Course 11 student in the way she didn't expand on or develop her response past a page of writing and on the way she treated Ambrose's text largely as content or information, which Theresa has seen both Course 10 and Course 11 students do. However, Theresa said that with Course 11 students, "there would be more of an essay. Maybe a couple of paragraphs. There might be some quotation which she doesn't have" (Theresa, 1010-04-11, Diagnostic Session). In response, Theresa decided the student was a fit to Course 10, although this student went on to jump to Course 12 at the end of the semester with Theresa noting that the student "worked hard" (Theresa, Course Results Sheet). Another student, 1010-01-07 was flagged as a Course 11A student

by both readers during the redo session, and Daria, upon reading her diagnostic, also ranked her as slightly high for the course, noting that she “might go to [Course 12] if she’s able to generalize less and use text to her own argument’s advantage” (Diagnostic Session). The student did not go on to jump, so perhaps she struggled to use the text as Daria indicated. Her end of course comment suggested that she “is growing as a reader and writer and should do well in Course 11” (Daria, Course Results Sheet).

Another seven problematic Course 10 students were flagged as higher than average students on a key criterion or behavior even though the course teacher did not flag them as slightly high overall. Student 1010-06-07 was flagged as an old Course 11A student by Adele during the redo session as he had “interesting” reading, but didn’t show “enough” work with text for her to be sure of what he was trying to say about the pieces (Adele, Redo Session 1). Daria, while looking over his diagnostic, also noted that he had “interesting thinking” in regards to his reading. In fact, she had some difficulty deciding where to rank him, but decided he had “high end reading” but needed to learn “academic forms and formats” so was still prototypical (Daria, Diagnostic Session). In her final note at the end of the course, Daria indicated that he was a “strong reader who never did the work to demonstrate this ability in writing” (Daria, Course Results Sheet). With student 1010-05-02, Theresa flagged the student as Course 11A during the redo sessions, and even though she still pushed the student down since there were some reading issues with “making things up,” she felt the student improved across the essay and the student’s reading got better when they started working with Pluto (Theresa, Redo Session 2). Adele, the Course 10 placement reader, saw the student as reading more fully in negative terms as to her the student “made [Pluto] fit their reading” of Graff (Adele, Redo Session

2). Hannah, the student's teacher, noted both tendencies as she saw the student "is spending more time focusing on" ideas of interest to him in the reading and "sidestepping" the actual task of the diagnostic, but he understood a good deal of Ambrose's text (Diagnostic Session). She put him as a "prototype of the course but higher end" (Diagnostic Session).

Finally, three additional problematic Course 10 students jumped straight to Course 12 at the end of the semester. For example, 1010-04-13 was flagged as Course 11A by both placement readers and was tweened in the redo session (although ultimately the same decision was made to place the student into Course 10). However, she was seen as prototypical of the course by Theresa on the new piece generated in response to the diagnostic. The student spent a good deal of the essay off task talking about "not sitting on the side" as a "soccer player" before attempting to connect Graff and Birkenstein to Ambrose (Diagnostic Session). When the student did try the assignment task, she "match[ed]" the "content," rather than the "concepts," which suggested to Theresa that the student doesn't think about reading as concepts but information (Theresa, Diagnostic Session). Yet, the student was described in Theresa's end of semester comment as a "class leader" and someone who "really worked on reading" (Theresa, Course Results Sheet) With student 1010-07-07, I previously hypothesized that Theresa likely made a different placement decision, and potentially an incorrect one, during the redo session because she was rushing to take a break from the sweltering room we were in. She said little about the student overall comparatively to other students, but claimed there was "very smart" thinking and reading in the piece (Theresa, Redo Session 2). As Adele read the student's diagnostic in her course, she ranked her as prototypical because she was

unsure “how much of this is just her and how much she’s getting from the reading” but felt the student definitely had some misreading at times (Adele, Diagnostic Session). Despite this, Adele said the student “could probably like do quite well though” (Diagnostic Session). By the end of the semester, the student jumped to Course 12 because she “used other texts to say something new, original” (Adele, Course Results Sheet).

In total, 18 of the 30 problematic Course 10 students (60%) show clear course designations or outcomes that relate directly to placement designations. For each of these students, the placement and course readings uncovered and noted similar or related strengths that both the placement and course readers found important to the course whether or not the students went on to use or develop those strengths. Further, when all 30 problematic Course 10 students’ grades are examined and compared to the 26 unproblematic students, the Course 10 students with problematic placement outperform their peers. First, they have fewer failing grades with 3.33% of problematically placed students failing to 15.38% of those students easily placed in Course 10. They also have a higher rate of jumping Course 11 and going to Course 12 with 20% of the problematically placed students jumping to just 15.38% of the other students. Finally, the problematically placed Course 10 students’ median grade, at a B, was higher than those easily placed in Course 10, at a B-, and was higher than the median grade of a B- for all Course 10 students that semester.

While there is a clear relationship between placement designations and course concerns and outcomes, there are a few students who raise a key concern with student motivation and Course 10 when examined more closely. Student 1010-01-11 was

designated as a Course 11A student in both sessions and had a different placement decision made about his essay in the redo session. When he arrived in Course 10, he really “wanted to be moved up,” as Daria said after reading his essay (Diagnostic Session). Daria was concerned that he had a tendency to “fall back on what he believes” and read to clichés rather than closely, so she kept him in the course (Diagnostic Session). Despite all the signals that suggested he was definitely strong in the course, including two placement designations and Daria’s own ranking, the student did not jump, received a B, and Daria said, “still working to strengthen his reading—essays will get stronger when that does” as a final comment on his work (Daria, Course Results Sheet).

A similar situation happens with students 1010-05-03, who was flagged as Course 11A in both sessions, who was tweened in the redo session where both readers decided to place the student up to Course 11, and who was ranked as slightly above by her Course 10 teacher, Hannah. However, the student ends up withdrawing from the course sometime between weeks two and five, a decision we can only speculate on, but which stands out as odd given the strengths the student brought to the class. Finally, student 1010-07-03 was also flagged as a Course 11A student in both sessions and was tweened with readers deciding a Course 11 placement in the redo session. On this last student, the final comment offered by Adele on the student’s final course work is potentially telling. Adele described a student who was “confident in terms of what [she] had to say,” but who “resisted writing a project,” and only did “minimal close reading,” and Adele concluded that the lack of close reading and “resistance to revising [were] holding her back” (Course Results Sheet).

While I cannot say what the best decision was for these students—either in terms of placement or the course—with only the evidence presented here, and I am certainly aware of students of my own who come in with promise that they do not develop, as is the determination of the teacher’s comments on two of these students, I do think that some Course 10 students resist if they feel they are overprepared in comparison to their classmates. And, for these students the possibility of jumping over Course 12 is not enough motivation, particularly when they feel the person who will make that determination is not reading them well already. Thus, they may only put in minimal effort rather than the full effort they can. It is necessary for Cardinal College to consider the demotivating effect of Course 10, even with the ability to jump over Course 11, for some students, especially as it decides what to do with old Course 11A students who are overprepared for Course 10 in some ways.

Problematic Course 11

Course 11 teachers, as we saw when comparing the criteria of the diagnostic sessions to those of placement overall, do not all share the same concern for reading. A look at students’ records across all data sets more closely reiterates this finding. Only one participating problematic Course 11 student was designated as slightly low by a Course 11 teacher. This student, 1011-09-06 was flagged in both sessions as a Course 11A student, and in the redo session, both readers indicated that label when the student was tweened. In this session, the placement readers decided together the same decision as the original session—that the student belonged in Course 11 because she did well at the end with “using what Graff is saying to interpret what’s going on with Pluto” even though where she started was “a mess” indicative of Course 11A students (Redo Session 2). As

Janet read the student's diagnostic, she noted the student was "a little bit below" the course because the essay was "very short, not understanding the task at all, and off the subject" even as it "[tried] to make a point" and had "a nascent kind of thinking" (Janet, Diagnostic Session).

Despite having this one course designation align with a placement designation, Course 11 teachers do take note of the issues that placement readers took note of, they just do not consider these issues to be obstacles to course performance. Fourteen out of the 24 problematic Course 11 students were discussed by course teachers for the same or very similar reading and writing issues as they were noted on by placement readers. Yet none of these students were flagged as slightly below the course by their teachers as placement readers indicated. This difference aligns with what was shown in the previous section when there was a misalignment of Course 11 diagnostic criteria and placement criteria. Course 11 teachers believe the course is meant to develop reading and writing abilities through work on close reading, and as discussed earlier, Course 11 teachers do not break down reading issues into subcategories of specific kinds of reading behavior, which may mean issues with reading are considered broadly rather than differentiated.

Despite this earlier finding, some student examples support Course 11 teachers' determination that students will develop the skills they need, while others suggest some students may not. For students like student 1011-03-06, there is ultimately success. She was designated as Course 11A in both sessions because of some misreading issues and was designated as such by both readers in the redo session when she was tweened, but she was placed in Course 11. She earned a B- in Course 11, and Odysseus noted "adequate reading; considerable improvement over the course of the semester"

(Odysseus, Course Results Sheet). Another student who did well, 1011-07-01, was flagged as Course 11A in the original placement and had a habit of writing “summary” (Theresa, Redo Session X) and “retelling” (Theresa, Diagnostic Session) versus working with the text. Yet, he “got beyond retelling” and had “good class work,” earning him a B (Theresa, Course Results Sheet).

Yet for students like 1011-02-07 who earned a C-, something was amiss in his performance. It is hard to determine what given that the end of course comment does not specify specific criteria, only that the student “did not revise papers for final portfolio” (Terence, Course Results Sheet). Yet, given the student’s earlier placement difficulties—he was designated as Course 11A in both sessions, was tweened in the redo sessions, and had his decision changed to Course 10 on a second reading—plus the discussion of his difficulties with reading at placement and at the diagnostic, it seems possible that what needed to be revised involved reading or working with text. Something similar happens with 1011-02-03, who was flagged at placement and in the course for reading issues and who earned a C- for not doing “content revision in either paper, only mechanics,” (Terence, Course Results Sheet). Another student also has a negative course outcome related to reading. Student 1011-09-01 was placed both times in Course 11, but was discussed both times as possibly “misreading” or possibly needing “to be told to pay attention to the text” and “unpack” (Theresa, Original Session 6). After failing to turn in a final portfolio, he earned an F. Again, while I do not know the specifics of any of these students’ cases from the perspective of teaching them, it is possible, given their earlier issues with reading, that they could have struggled to do the things required of close readers in Course 11. If their reading issues caused their struggle, and not say effort,

attendance, or the like, they either needed to be placed in Course 10 or to have different support in Course 11.

There is no doubt that some of the problematic Course 11 students go on to successfully pass the class, and so perhaps they are not concerning after all. But enough go on to struggle that problematic Course 11 students taken as a whole underperform in relation to their unproblematic peers in terms of final grades. When all 26 problematic Course 11 students' grades are examined and compared to the 25 unproblematic students, the Course 11 students with problematic placement underperform in relation to their peers. First, they have more failing grades with 7.69% of problematically placed students failing to 0% of those students easily placed in Course 11. Also, the problematically placed Course 11 students' median grade, at a B, was lower than those easily placed in Course 11, at a B+. However, at a B median grade, the problematic Course 11 students matched the median grade for all students that semester. A key difference though was the number of students who fell below the median grade of Course 11 overall that semester rather than at it or above. 30.77% of problematic Course 11 students fell below the median grade of a B, but only 8% of unproblematic Course 11 students fell below the median grade of a B. These differences, differences that show problematic students as largely underperforming in relation to their unproblematic peers shows that something noted at placement is relevant to the course outcome.

Further, the program must consider the cumulative effect of underperforming on reading and writing, particularly as the difficulty level only increases as student's move to Course 12. Will these students, who pass Course 11 although behind their peers, go on to face a widening gap when the tasks are more complex and there is comparatively less

support? Will the student who writes “summaries that are phrased as arguments” be able to hack it when she finds herself in a sea of research and must build her own research project on something bigger than information? (1011-09-06, Janet, Course Results Sheet)

Findings on Link Between Placement and Course Concerns and Decisions

To sum up, Course 10 teachers’ concerns are aligned quite well with the placement process and system as it already exists. There are a good number of shared designations across placement and the course, and course outcomes are related to determinations made during placement. For Course 11, the same alignment is not seen. While readers note similar characteristics of student’s work, the characteristics—particularly in regards to reading—are not concerning to all Course 11 teachers. In short, the reading distinctions made during placement do not appear to carry the same weight with course teachers in Course 11 as imagined by placement readers. However, the course outcomes, in terms of grades overall, do show some relationship between placement designations and course outcomes. As the writing program at Cardinal College decides how to treat old Course 11A students more systematically, it must also determine what weight reading will play in placement—specifically which reading behaviors are most concerning at placement—and how it will balance the possibility for over prepared students in Course 10 to feel demotivated against the possibility for underprepared Course 11 students to potentially struggle in the course, even if only slightly, and go on to potentially struggle more when they reach Course 12.

In response to these concerns, several recommendations will be sketched in detail in Chapter Five. First, the program must have shared discussions on the relationship of reading to writing and what this means for the teaching of our particular student

population so that Course 11A understandings of our students are not lost and that all Course 11 teachers have the opportunity to benefit from another way of understanding students' reading habits and behaviors. These discussions can allow the program to talk together about resetting the boundary of these courses given the now permanent loss of Course 11A. It also would allow teachers to design and share strategies for responding to and addressing reading difficulties specifically, as part of staff development, potentially making an impact on the performance of the most concerning students in the program before they move on to Course 12, where they will conduct their own research—finding, reading, and building on the work of complex academic texts—without shared readings that provide scaffolded reading support.

Second, a revision to the placement reading method is suggested where the prime directive of Smith's (1992/2009, 1993) expert-reader model, which asks teachers to use their teaching experience, is reasserted, but where concerning students—those a reader feels torn about—are tabled until both readers can make a decision together (Haswell & Wyche-Smith, 1994), sharing and weighing their different concerns in conversation. Such a process would allow the program to suspend or bracket out their old labels and consider problematic placements in a number of ways before making the ultimate decision. This way, the program could ensure old Course 11A students are considered from all sides, it could collect valuable student examples for subsequent staff development, and it could develop ways of tracking these students for future study. Going forward, future study of these students would be necessary to determine the impact of any specific course and teaching changes from the departmental conversations above, as well as from changes to the placement reading procedure.

Conclusion

The study found that placement readers at Cardinal College see students as existing on a continuum of ability linking reading and writing, as the program asserts, and they indeed use teaching specific criteria to determine a student's position on that continuum, presenting a pedagogically useful view of students' placement essays. The placement criteria, which focus primarily on how students come to and represent a reading of complex texts through writing, expose students who, past experience and the current research shows, may go on to struggle in a mainstream college writing course because of particular combinations of reading and writing behavior. However, the study also found that because of the change to Cardinal College's curriculum, with the loss of a four-credit course between the two remaining courses, the expert-reader model (Smith 1992/2009, 1993) that placement readers previously operated from is no longer being followed as the Course 11 reader splits her focus between her recent Course 11 teaching experience and her long-standing knowledge and expertise on students in the eliminated course.

As the Course 11 reader recognizes behaviors of students in the eliminated course, she stops making a decision only about Course 11, and instead weighs the merits of both possible existing course placement options for herself, often making a notation on a student's file suggesting the advising office register the student in the section of a particular teacher who has experience teaching the old course. In short, the expertise of the Course 10 reader is not being fully engaged to help discern which students would be most or least concerning. And the current practice of recommending troubling students to only some teachers may exacerbate differences between two subsets of Course 11

teachers and is not a long-term solution. Yet, the anxiety of the Course 11 reader is not unfounded. Pushing students down to the six-credit course option, when it may not be necessary for all students, could have consequences for students' progress through their courses of study and their engagement with the learning in the writing program, not to mention cause pushback from students, other departments, and the administration. Further, an analysis of the course criteria shows that Course 11 teachers with different past teaching experiences—either Course 11 or old Course 11A—do have different understandings of students' reading difficulties that come up at placement and could potentially affect how they respond to students' reading issues.

On either side, Cardinal College faces the possibility of a good number of inadequate placement decisions, as the data have shown that the majority of the problematic students studied here are truly atypical in terms of criteria and could go on to perform differently in the course than their teachers may have anticipated, which is particularly important for problematic Course 11 students who could underperform in Course 11 and go on to struggle further in Course 12 without proper curricular intervention, but who could also feel frustrated in Course 10 if it is not a sufficient challenge, causing a lack of engagement with academic reading and writing.

It is the conclusion of this study that Smith's (1992/2009, 1993) expert-reader model must be reasserted, with a modification made by Haswell and Wyche-Smith (1994) that has readers not make a placement decision on students who do not fit neatly into their existing course or on who they have reservations. I argue that each reader must wait to make decisions on problematic students until they can do so in concert with the other reader through conversation. While this model will likely extend reading time, it

will engage the expertise of both readers and produce conversations about troubling student work that the writing program can use to redraw the lines between these courses. And, the details of these conversations and rich student examples can feed departmental conversations about the work of these courses and prompt staff development workshops where teachers share strategies for addressing students' needs.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

At the start of Chapter One, I exposed the concern that is at the heart of this study and a key question of writing assessment, and I'd like to return to it now: How do educators know that we are making the best decisions for students? The short answer is that we have to investigate, considering relevant evidence to have some assurance—some validation—that adequate decisions are being made for students. While this short answer is spot on, it hides a great deal of conflict that has arisen out of a historical separation of the work of teaching writing and the work of writing assessment. Since the early work of writing assessment had been developed primarily by researchers working in the field of testing or psychometrics, and only later contributed to by writing teachers, the two different communities and their conflicting concerns, focuses, values, and conceptions of writing (Williamson, 1993) have been at odds.

A Brief History of a Divided Field for Writing Assessment

Across the history of writing assessment, a seeming war (White, 1993) has been waged as teachers fought for ways of assessing writing that adequately addressed the reality of writing and the needs of teachers and learners. However, looking at the early history of writing assessment (1870s to 1960s), teachers' concerns went unheeded and theories of writing were seldom reflected in early assessments. Instead, the commitment of many test specialists and psychometricians was to considerations of cost, efficiency, and objectivity (which, at the time, was understood as requiring decontextualized assessments) since these considerations were necessary to meet the growing demand for tests that could be administered to large numbers of people (Elliot, 2005). Testing

agencies like the College Board and ETS focused on creating large-scale, standardized assessments with the goal of assessing and measuring ability regardless of whether it was the ability of U.S. Army personnel for government decisions of recruit placement or the abilities of an increasing population of college students for educative purposes (Elliot, 2005).

The reality and emphasis on large-scale testing concerns—coupled with the cultural belief that a standardized science of testing could be wielded to fairly and systematically determine the ability and merit of all—made it difficult for test creators to hear the writing theory and teaching-based concerns that were continually raised by English teachers (Elliot, 2005). Such concerns questioned if fair and instructionally significant decisions could be made by early writing assessments that did not actually collect students' writing, that knew nothing of the specific details of writing theory, and that had no regard to the particular contexts of teaching. Yet, even after writing teachers like White (1973) became involved in writing assessment, the work typically still stressed method and practice over theory and still relied heavily on testing concepts (like reliability) that were part of a problematic tradition that was not congruent with the teaching of writing (Huot, 1996/2009).

Reconnecting Assessment to Teaching and Learning: Validation

Seeking to fill the gap, and to argue against past assessment practices that did not align theoretically with the teaching and theory of writing, scholars like Huot (1996/2009, 2002) and Huot and O'Neill (2009)—as well as the Conference on College Composition and Communication's (CCCC) influential policy document *Writing Assessment: A Position Statement* (2009)—have argued for and begun developing theory to guide

ongoing writing assessment work. One major focus in this work has been on reestablishing the primary purpose of assessments: to support teaching and learning goals by informing and enriching teaching and learning. Current theory claims that writing assessment should, among other things, be site-based and locally-controlled (Huot, 1996/2009) with pedagogical concerns playing a key role in how writing assessments are developed and refined (CCCC, 2009).

Additionally though, as Huot (2002) sought to facilitate the development of local assessments, he argued that newer theories of validity in educational measurement would actually provide writing teachers a structure for validating and for making a case for their local writing assessments to various outside stakeholders—both testing specialists and college administrators. These theories would be useful, he explained, because rather than presenting obstacles for writing teachers, as many prior versions of such theories had, newer theories had begun to recognize the “social construction of knowledge” (Huot, 1996/2009, p. 161) which aligned testing theory with writing theory and provided writing teachers with a way to make testing theory useful.

While terms like validity and reliability come out of a psychometric tradition that once stressed large-scale, objective, one-size-fits-all tests over context specific assessment, such terms have not remained static. As the field of educational measurement developed, the theories behind these terms were revised. Newer theories of validity stress the need for ongoing study into an assessment process with the goal of crafting an argument for the adequacy and appropriateness of the use of an assessment, including thought given to the alternative arguments, the decisions to be made using that assessment, and the consequences of its use (Cronbach 1988; Kane, 2006; Messick,

1989a, 1989b). It is these newer theories that Huot (2002) argued could be useful for investigating and explaining our local writing assessments since they required that the particular use of an assessment in a particular context is what was being investigated in validation. Additionally, these newer theories make ongoing validation part of the responsibility of test users and creators who must now consider alternative explanations and the challenges of their most critical stakeholders (Cronbach, 1971).

The theory or understanding from which a writing program operates—both in terms of their courses and their placement process—becomes an argument they make for the adequacy of their placement decisions. And it is this very argument that must be investigated in validation research. Validation research is already being carried out by specific writing programs interested in investigating their placement systems, and just as Huot (2002) suggested, the use of educational measurement theories has allowed these writing programs to investigate their arguments and reconnect assessment to their teaching and learning goals (see, for instance, studies like Gere, Aull, Green, & Porter, 2010; Gere, Aull, Perales Escudero, Lancaster, & Lei, 2013; Elliot, Deess, Rudiny, & Joshi, 2012). It was a desire for assessment to be usefulness for teaching and learning that attracted Cardinal College's writing program to validation research. Cardinal College sought to make use of Huot's (2002) suggestion that newer validity theories are useful for responsibly checking the assumptions of a writing program, addressing the concerns of its interested stakeholders, and focusing on teaching and learning by beginning validation work into their placement process, and in that way adding to recent validation literature.

Why Was Validation Useful to Cardinal College?

Cardinal College had an emerging pair of concerns not unlike those sketched in the history of writing assessment above that could and should be addressed through ongoing validation research. First, we had been operating on the assumption (and had made the argument to administrators) that we use course criteria to make course relevant placement decisions, and we wanted to explore that assumption, as it was important to have data to support the assertions of our interpretive argument and to address the push from the administration for the implementation of an indirect test that would have faster score turnaround.

Some years ago (around 1998), Cardinal College adopted Smith's (1992/2009, 1993) *expert-reading model* to make placement decisions. Through Smith's own investigation into his placement process, he determined that the most adequate and reliable placement decisions were made by teacher/readers who focused on the course they had most recently taught and decided if a student belonged in that course or did not. Since the adoption of Smith's model, the writing program has argued that because it uses placement readers who have most recently taught particular courses, it is making course-related placement decisions. Cardinal College has asserted that expert placement readers tap their teaching knowledge and experience to make more adequate placement decisions since those decisions are based on a locally-developed and direct writing assessment that is read using criteria that align with the writing program courses.

Drawing on Smith's (1992/2009, 1993) model, Cardinal College's writing program has pushed back against the administration's suggestion to adopt a nationally-developed placement test, choosing instead to assert its locally-developed test and the

expertise of its teacher-readers. Yet, Smith's model was based on formal study into his placement process, and such work has never been done at Cardinal College. Therefore, validation research would help Cardinal College investigate its assumptions and either substantiate its claim or better perfect its placement process to meet its goals.

Second, several years ago (around 2008), a four-credit course that fell between the two remaining courses in the program was eliminated due to space constraints on campus that caused the administration to regulate credit hours, fitting courses into consistent time slots for scheduling ease. With the loss of this course that fell between the two other course options, the writing teachers have been concerned about the potential for students to not fit as neatly into the remaining course options. The courses were crafted to be developmentally connected and imagined for Cardinal College's particular student population, who display a diverse range of academic reading and writing abilities. The writing program has worried if the loss of the middle course, which disturbed the intended boundaries of the courses, has had a negative effect on the courses and placement into them. On this issue again, through formal study, with an emphasis on evidence gathering and investigation, validation research can help Cardinal College study this concern.

Study Focus and Research Question

Given the above concerns of Cardinal College's writing program, the current study was developed. The focus of this study overall was a beginning inquiry into the validity of the placement process at Cardinal College (a pseudonym)—a small, private college in the suburbs of a mid-sized, Mid-Atlantic city. Cardinal College's writing program uses Smith's (1992/2009, 1993) expert reader model to place students in one of

two reading and writing integrated courses. Recall, in Smith's model, an expert rater is one who has recently taught the course for which they are making placement decisions, and in his findings, these raters more reliably and adequately make placement decisions by employing their teaching and learning expertise to look not only at students' writing, but at the undertaking of teaching these students. Given this focus, this study considered the following research question:

- Do expert readers at Cardinal College use curriculum/course related and teaching specific criteria to offer a pedagogically useful and purposeful, view of students' placement essays? Are expert readers' criteria (and by extension their decisions) linked to course criteria?

To answer this question, it was necessary to break the research question into four subquestions:

1. What are expert-readers' placement criteria?
2. Are expert-readers' placement criteria teaching specific?
3. Are expert-readers' placement criteria linked to course criteria?
4. Are expert-readers' placement concerns and decisions linked to course concerns and decisions?

Pursuing its research question and subquestions, this study began the process of crafting a *validity argument* for Cardinal College's placement by specifying the program's *interpretive argument* (Kane, 2006, p. 22) and investigating one key assumption of the writing program's argument—that placement readers place students according to course content criteria. Of course, this one study is not the whole of what is required for

validation of Cardinal College's placement process. Future studies would have to investigate other assumptions, as well as insight gained from this study.

Study Design and Participants

Cardinal College's interpretive argument specified, among other things, that placement readers, who are expert teachers in the writing program with recent course-teaching experience, use their knowledge and experience of the courses and the students in the program, their *conceptual frameworks* (Kane, 2006), to make placement decisions directly on students' impromptu essays written in response to a prompt that asks students to use one provided text to analyze the scenario presented in another. The research question and subquestions of this study were investigated in a number of ways through multiple data sets.

By conducting interviews (Denzin & Lincoln, 2008) and think-aloud protocol sessions (Ericsson & Simon, 1984) with two placement readers while they read placements and with seven course teachers while they read student diagnostic essays and course work, the assumption that Cardinal College uses course-related criteria in placement was examined. Through analysis of these interviews and the think-aloud sessions where readers and teachers discussed a total of 119 participating students' work (49.17% of the total number of students enrolled in Course 10 and Course 11 for the semester of study), the study uncovered common and recurring criteria that emerged in the placement of each of the two courses at the College—Course 10 and Course 11.

Using these criteria, a relationship was explored between placement and teachers' course criteria that offered one data set relevant to the research question. Additionally, the research question was also examined through teachers' impressions and course

grades. As the course teachers completed a survey tallying the number of students overall who they felt did or did not fit well into their course, as they detailed their impressions on the adequacy of placement during think-aloud sessions on diagnostic essays and in interviews, and as they reported final grades and assessments, the teachers offered their impressions of the relationship of placement decisions to course work.

Summary of the Study Findings

The analysis of the placement criteria and the concerns of readers revealed first that Cardinal College's placement readers see students on a continuum of linked reading and writing ability, just as the program and courses posit. Second, the exact border between the two courses, Course 10 and Course 11, was impacted by a reader's sense of the potential for a student to experience reading difficulty. Third, that potential was hard for readers to determine when the single piece of student's placement writing had characteristics that provided insufficient evidence for the reader to feel confident about their determination on the strengths and limits of a student's reading ability.

Thus, the readers frequently used teaching specific knowledge expressing pedagogical concerns centered on how to teach students with particular reading and writing behaviors to determine where problematic students might struggle or excel, as they should given Smith's model. Yet, this knowledge often engaged readers' previous experiences teaching students in a now discontinued course that fell between the two existing courses, Course 11A. As a result, Smith's expert-reader model was not being properly followed as the Course 11 reader worried about the potential for misplacement given the loss of the old course, and her attention was divided between considering both course options for old Course 11A student essays rather than making a decision for one

course. To address her worry, the reader often placed old course students with teachers who had experience teaching the old course, and this decision has the potential to exacerbate existing differences in teachers' understandings of the student population.

Furthermore, the analysis of the course criteria and the concerns of course teachers revealed that Course 11 teachers were divided into two groups based on their beliefs about the necessity of assessing the incoming reading abilities of Course 11 students, and that divide coincided with previous Course 11 teaching experience. Those who once taught Course 11A attempted to discern how much and in what ways a student would struggle to read the kind of texts in the course, and those who hadn't primarily taught Course 11A believed incoming students would struggle in similar ways addressed already by the course. Therefore, placement designations were not aligned with all Course 11 teachers. However, Course 10 teachers, in contrast, had a shared set of criteria and vision for their course, and a strong sense of the specific reading habits of underprepared students valuable for distinguishing less severe reading habits and behaviors from more severe ones, which aligns with placement decisions made regarding Course 10.

Finally, an analysis of all data, including teachers ranking of students' fit to the course, as well as final grades and course outcomes confirmed the earlier finding that Course 10 teachers' designations and decisions aligned with placement ones while Course 11 teachers' designations and decisions were not fully aligned. However, the course outcomes showed that there must be some relationship between Course 11 placement decisions and Course 11 since students flagged at placement as old Course 11A students underperformed in relation to their Course 11 peers in terms of course

grades. Finally, since Course 11A students also outperformed their Course 10 peers, the program must consider the negative consequences old Course 11A students could experience in either placement, as students face demotivation on one end and the potential for insurmountable difficulty on the other.

Conclusions and Discussion

From the findings of the study, several conclusions can be drawn. First, the biggest current impediment to the assumption on which the placement process is built—that placement readers use course-relevant criteria—is the loss of Course 11A, a course that once existed between the two remaining courses. This impediment is supported by Smith (1992/2009, 1993) who discussed course boundary issues, where different readers have different imagined boundaries for the courses, as being a possible contributing factor to reader disagreements. Because Cardinal College sees students as existing on a continuum of academic literacy acquisition, the loss of this course means that placement readers in the program do not feel as confident in their ability to discern students in the middle of the continuum, and by extension they do not feel as confident in their decisions.

Readers once placed students on either end of the continuum into the two existing courses—those with severe reading and writing difficulties in Course 10 and those with some existing knowledge of and abilities with academic reading and academic forms in Course 11. Course 11A, for placement readers, was for students who fell toward the middle of the continuum. They showed either a mix of some reading and writing difficulties, or were more severe on one or the other side, either reading or writing, but through more directed reading instruction, they could be caught up to their peers without

requiring a six-credit course. These old imagined boundaries have been altered, and that has affected confidence, but another factor seems to affect readers' confidence as well: the consequences of inadequate placement decisions.

Readers worry about what will happen to Course 11A students as they are placed into courses designed for students on either end of the continuum rather than for these students who fall more toward the middle. If Course 11A students are placed low, they would be burdened by a six credit reading and writing class, potentially unnecessarily, and such burdens can have serious consequences for student retention, engagement, and growth. Yet, if they are placed high, they could struggle and fail to keep pace with their peers, either in Course 11 or as they move to the second required course in the college writing sequence, Course 12. In short, the lack of confidence is about not being able to meet the goal of placement testing, which as Kane (2006) describes, is choosing a course that is optimal for the student's current learning level.

Thus, a second conclusion is that while readers in the program are not making fully consistent decisions, and while a few of their decisions can even be debated in terms of their adequacy, it is a deep concern with placement adequacy that currently disturbs the Cardinal College placement process. A teacher's conceptual framework (Kane, 2006) includes the context of the work of specific courses and experiences with a particular student population. When a teacher-reader's conceptual framework suggests that existing course options, as they are conceived, are inadequate for specific students in the population, how is that teacher-reader to make adequate placement decisions? In the case of Cardinal College, the solution has been to unofficially continue placing students into the now eliminated Course 11A by recommending that students who show old Course

11A attributes be placed into sections of Course 11 taught by teachers who have past Course 11A experience. Thus, Cardinal College solves a concern with placement adequacy by trying to maintain the eliminated course, even unofficially, so readers can feel more confident that students' needs are being addressed.

As this solution addresses fears of placement adequacy, it also causes potential adequacy problems, both for placement and teaching, as I've discussed in Chapter Four. For Cardinal College's placement readers, the weight of the decision to be made for old Course 11A students—i.e., their concern with placement adequacy—calls to mind their old course expertise. It is this old expertise, built inside a context that no longer exists, that is tapped during placement. Further, despite arising from a concern with adequacy, the decision to place old Course 11A students with old Course 11A teachers who are teaching Course 11 means that Course 11 runs with two different sets of expertise and two different populations. The residue of this old expertise likely prevents new teaching and placement expertise from being built inside the new course context.

A final conclusion drawn from these findings is that some action or actions must be taken by the program as the loss of Course 11A is now officially permanent and the difficulty placing old Course 11A students means the program does not have a structural approach that accounts for these students and their particular needs. Whatever action is taken, it must work to address these students' needs, particularly in regards to reading, if the courses and placement decisions are to be more strongly aligned. As I continue, I detail three recommended actions for Cardinal College's writing program going forward.

Recommendations

Building on these conclusions, which show both a deep concern with placement adequacy and an uncertainty about how the program can make adequate placement decisions after the curricular change, this study makes three recommendations for Cardinal College's writing program: one in regards to a new placement reading process, one in regards to teaching the courses as they are in flux, and a final one about departmental work reimagining the courses or their boundaries. While the major concern for the program is probably the work of reconceiving the courses, I will discuss this recommendation last. I believe the work of reimagining the courses will take both time and discussion, and the work of placing and teaching students must carry on in the interim. Further, I think the discussion of the writing program staff will benefit from what readers learn through the new placement reading process. As I continue, I detail each recommendation, providing the necessary rationales and suggestions for implementation.

New Placement Reading Process

The first recommendation of this study relates to the placement reading process. The problems uncovered in the placement process stem from the loss of Course 11A and the uncertainty about what will happen to these students inside two courses that were not originally intended for their needs. Thus, the biggest flaw in the placement reading process is that readers often attempt to make a placement decision by themselves about students who are not a neat fit to their course, and they are reluctant to push these particular students down or up. The Course 11 reader in particular is likely afraid because she reads first. It is possible that the Course 11 reader could be worried that the Course 10 reader will receive a set of essays with the knowledge that they have already been pushed

down by the Course 11 reader. The Course 11 reader is afraid her decision to push a student down, which comes to the Course 10 reader with no notes or comments, hides the complexity of the Course 11 reader's reservations and concerns about the student's writing and could potentially be read by the Course 10 reader as an endorsement of a Course 10 decision.

It is the recommendation of this study that Cardinal College's placement reading process must be adjusted. First, Smith's expert-reader model must be reasserted as the ultimate goal, but since course expertise is in flux, the placement reading process must be modified to account for that. A modification comes in the form of employing more of the *two-stage reading procedure* discussed by Haswell and Wyche-Smith (1994). Haswell and Wyche-Smith had initial placement readers, less experienced teachers, read and place only those students who they could easily decide belonged in the college level course in the span of a minute or so of reading. During the second stage, the unplaced essays were turned over to more experienced readers—experts on the courses—and were subjected to “as much study and discussion as needed” (Haswell & Wyche-Smith, 1994, p. 229). A large part of Haswell & Wyche-Smith's reasoning for deciding on this reading procedure relates to the fact that they were directly developing a placement system that had not yet existed, but that facet of their design could work for Cardinal College's particular situation, where placement decisions have become more difficult in the face of the lost mid-level course.

What I recommend is that expert placement readers still read during both stages, but that they place only those essays they can easily decide on within a few minutes' time in the first wave of reading. This process would require that readers admit reservations

and put aside student essays about which they have concerns. Then, in the second stage of reading, the same readers will make decisions on the unplaced essays, but now through discussion with each other. This process will engage the expertise of the Course 10 reader, which has been shown to be relatively stable, and produce conversations about troubling students. It is my belief that such a process can help readers feel less nervous about the consequences of placement decisions for difficult essays, as their worries and concerns—including their mixed sense of the students' abilities—can be heard by another person and discussed. Some of the most powerful moments of placement discussion in the data collected for this study took place during reading sessions focused on making a decision on *tweeners*—those pushed down by one reader and back up by the other (Smith 1992/2009, 1993). In these sessions, readers came together, shared evidence, and made a placement decision in conversation. Sometimes both readers found they shared the same mix of concerns, and they weighed the evidence together; other times one reader's view was changed by considering a new way of looking at the attributes of the student's essay. Regardless, these conversations led readers to feel that what they saw in a student's essay was considered, and that feeling is precisely the solution to the uncertainty that plagues the placement process currently.

Further, there is support in the literature that a focus on disagreements can lead to fruitful conversations about writing that develop expertise and improve teaching and learning (Barritt, Stock, & Clark, 1986; Stock & Robinson, 1987; White, 1993).

Therefore, as I move to outline the second and third recommendations, I want to say in closing here that the new reading process and the conversations it generates can be useful as the writing program enacts the other two recommendations that I will detail next. The

recommended reading process will uncover specific examples and instances of uncertainty, just as this study has already provided such examples for consideration. As placement readers read inside the new process, they will flag students and uncover and name common issues. Having these examples and issues will be useful for course teachers who are currently teaching the courses with these students, and they will be useful for departmental discussion and staff development reimagining the courses.

Teaching Through the Transition

The second recommendation of this study follows from the first. If Cardinal College placement readers must acknowledge and verbalize their reservations about particular students to bring to light the underserved population and solve problems of placement as part of a team of experts, so too must course teachers. Since individual expertise is potentially unstable as the program seeks to readjust the courses, the teachers in the program must also find a way forward through conversation and collaboration.

There are several ways teachers can acknowledge and discuss pedagogically relevant concerns as they teach through the transitional period in Cardinal College's courses. First, as teachers recheck placement during the first week of the course, they must overcome their usual reluctance to flag students as potentially misplaced (Smith 1992/2009). Rather than thinking they are definitely deciding that a student will be moved to the other course, the teachers must see their work as flagging student essays that would benefit from another look and a conversation with other course teachers. Second, groups of course teachers with mixed backgrounds—some having taught Course 11A and some not—should meet to discuss their diagnostics (and later even other course work) together.

The conversations between teachers can help the work of both placement and teaching. For placement, noting which students are still concerning even after a second piece of writing (or after a course paper) can help the program ask again if the best placement decision was made. Additionally, the program can better discern which students to track and follow throughout the semester (and even into the next semester) to gather more detailed information in relation to future placement decisions. As the teacher moves forward, they can follow up across the semester with troubling students flagged at placement and determine if their earlier concern was ultimately unnecessary or not. In short, they can ask themselves if the earlier noted issues continue to be a problem for the student, or if the course addressed these issues adequately.

But the placement benefit may be secondary to the teaching benefit. As teachers work together, they can draw on each other's differing experience to name and categorize specific concerns for specific students, and these are concerns the course teachers could continue to assess for problematic students going forward. What we come to understand about a particular student and the work of teaching that student is improved as we have more interactions and pieces of evidence, but also as we are instructed in different ways to interpret this evidence. For as Tierney and Pearson (1983) show about readers in general, reading is a construction of meaning, and developing as a skilled reader requires experience with enough texts of a certain kind and, frequently, challenge to our initial interpretations that help us make new ones.

Additionally, by meeting and working in groups of other teachers, each teacher will be presented with other ways to think about the course and particular students and student populations, and each teacher will have a network of other teachers from which to

draw support. Ongoing conversation and collaboration across the semester could help teachers plan and refine their teaching, work on and request staff development, and make decisions about students. For instance, teachers might work together to discuss how they address certain student issues through commenting styles, mini-lessons, or workshop. They might compile questions or concerns to request staff development or even develop presentations on certain questions or concerns for staff development. Teachers might work together to decide if particular students should be directed to the writing center for additional support, or even if they should be required to go to attend sessions. In short, while the program and courses are going through a transition, teachers in the program should turn to each other for assistance and support more frequently to make sure all student populations are fully seen and all teachers are taking note of course determinations that would be helpful for the reimagining of the courses.

Reimagine the Courses

Third, it is the recommendation of this study that Cardinal College's writing program must address more formally and structurally the loss of the old Course 11A by reimagining the work of the remaining courses to meet the needs of the Course 11A portion of the student population. The program tried already to secure permission from the administration to reinstate Course 11A, rather than readjust the courses, but the administration was unwilling to allow it, even when the course was proposed to run at time slots that would lessen the impact to space and scheduling constraints.

Given this reality, the only option is for the program to focus on reimagining the remaining courses without Course 11A, but accounting for this student population. Since not all teachers in the program have had many experiences with these students, at least

not in large numbers, the work of reimagining the courses will take some time and discussion. Future departmental meetings and staff development should focus first on teachers sharing their experiences and strategies for working with students from all perspectives—Course 10 teachers, Course 11 teachers, and Course 11 teachers who used to teach Course 11A. These initial discussions could draw on observations from the previously mentioned placement and course teaching changes, and these discussions would allow teachers who are less familiar with Course 11A students to share their experiences while having the opportunity to see these other students—the difficulties these students face, particularly their reading difficulties—as more than just idiosyncratic, but prototypical of a good number of students in our population.

As the discussion moves forward, the focus should shift to naming common student difficulties or behaviors, conceptualizing the thinking behind these behaviors, describing how students typically progress, and highlighting successful strategies for working with these students. Talking about the reality of teaching students with particular difficulties can help course teachers feel less alone or less unprepared to face these challenges. Further, the ultimate goal of this process is to reach a stage where decisions could begin to be made about where to redraw the course boundary lines. Departmental discussions drawing from actual placement and teaching examples will be the evidentiary base for a new programmatic vision that accounts for all student populations.

Since the placement process will have flagged certain students and their essays as examples, and since the teachers will have concerns about specific students and their learning, the entire writing staff can discuss the most problematic students flagged during placement or in the courses and try to come to a conclusion about how to handle

particular problems in different approaches to a revised curriculum and through different methods of instruction. A conversation across the department could tease out the most concerning habits of reading (or habits of reading in combination with writing), and what it takes realistically—in terms of commenting styles, sequencing of assignments, one-on-one work, persistence, etc.—to work with students on these issues.

It will be essential, as this conversation proceeds, to protect against two possibly dangerous routes of discussion. First, teachers who are unaware of the difficulties many first-generation college students face in acquiring academic discourse may assume that some students are just not college material. Yet, as Gee (2003) argued, not all students have had the same *opportunity to learn* the kinds of reading and writing expected in specific social contexts like the academic context of school. Thus Cardinal College, with its substantial population of first-generation college students who have likely not had this opportunity, cannot afford to build its program on the assumption that academic reading and writing that presents difficulty is a form of weeding out students who lack innate ability. Rather it must assume that ability is built as students are exposed to new and complex reading and writing tasks that exist inside the social context of the academy—that is, difficulty is not a sign that students cannot do college work, but rather that they likely need more interaction and experience with such work to properly acquire ability in the academic discourse.

Second, discussion that advocates for the courses to progress from skill-and-drill reading and writing tasks toward more complex work must be guarded against as well. It is a frequent logical misstep for some teachers, particularly those who primarily teach upper level courses, to assume that students who display reading and writing difficulties

must not have mastered prerequisite skills of a lower order, and by extension, the assumption goes that these students must master these prerequisites before they can graduate to higher academic work. In essence, this belief supposes that students must master the rules of grammar before moving to paragraphs and on to essays, or that students must work on basic text comprehension before taking on interpretation or analysis. The reality is that students who are asked to do more complex academic work will display a marked depreciation in other skills as they grapple with new difficulty, and that waiting to introduce students to the actual reading and writing work of college will not provide them opportunities to learn academic discourse. Conversation should instead focus on the particular difficulties that can and cannot realistically be addressed inside Course 11 while still preparing students to take on the demands of Course 12.

Future Studies

As I have already specified a number of times throughout this study, validation is an ongoing process (Kane, 2006). Cardinal College's writing program, and all programs, must continually study their work to ensure that their assessment decisions and processes and the assumptions behind those decisions are logical, consistent with what is happening, and to good effect. The work that this study undertook represents only a beginning to the work of validating Cardinal College's placement process, and since this study, while confirming some relationship between placement decisions and the course, also uncovered a disagreement about the value of the key placement criteria of reading to deciding placement, the program must continue to study the alignment of placement decisions to the course. In the following section, I detail two possibly useful studies for the program going forward.

Tracking Consequences to Course 11A Students

First, one of the issues uncovered when analyzing Course 11 students' final grades is that old Course 11A students as a group did not perform as their peers did. When these students were placed into Course 10, they did better than their classmates and when they were placed into Course 11, they underperformed in relation to their peers. This may seem like an endorsement for placing these problematic students into Course 10, but it is not. While Cardinal College's writing program does not consider or treat Course 10 as if it is remedial, students likely understand the course in these terms given that they know they must still take required courses in addition to Course 10 and they likely have friends who did not have to take Course 10. Furthermore, there are many students who show promise at the start of Course 10 who do not go on to perform as expected. It is true that our initial expectations for students can be incorrect, but it is also true that some students are demotivated when they feel over prepared for a course but unrecognized by the teacher for their abilities.

On the opposite end, there is another problem. Old Course 11A students who are placed in Course 11 face the possibility of struggle. Difficulty, in and of itself, is not a problem. In fact, part of what college students learn generally is how to face difficulty on their own, but difficulty must not be overwhelming, or it does not lead to learning or growth. Given the difference between problematic and unproblematic Course 11 students' final grades, it is safe to assume at least some of these students faced a great deal of difficulty, but we can't be sure exactly where, why, or what effect that difficulty had on students since this study did not collect student perceptions as part of its data.

I propose that a key future study conducted by Cardinal College's writing program should track the consequences of placement decisions for students, particularly those students who are not a neat fit to existing courses. Further, student perceptions must be collected as part of the data for this future study. Messick (1989a) expressed that the consequences of a decision should play a role in the determination of its validity, and as such Cardinal College must investigate the consequences of its placement decisions.

It is also important to note that the tracking of these students should likely carry on past the completion of Course 10 and Course 11. It is possible that students who underperform in Course 11 but pass will go on to experience deepening struggle in Course 12, as the cognitive load is greater once students must do more than just read and write in response to a teacher's prompt. In Course 12, students must create and shape an academic research project of their own—with their own goals, purpose, key materials, and questions. Cardinal College has an obligation to track how students perform in Course 12 as well, understanding the long reaching effects of initial placement decisions.

The Placement Prompt

As the interpretive argument for Cardinal College's placement process was sketched in Chapter Three, two assumptions were uncovered. The first, studied here, looked at the assumption that placement readers' criteria are course-related. A second key assumption of our placement process was that the placement prompt was intimately related to the reading and writing construct and tasks of the courses, and therefore assisted expert placement readers in making a course-related decision about placement. Investigating our assumptions about the placement prompt and its relation to the course

are important to validating Cardinal College's interpretive argument as well, and a future study should study the prompt.

In addition to considering how the prompt relates to the course, pieces of this future study should consider several concerns about the writing prompt raised by course teachers in their interviews for this study. One teacher raised a concern about different student and teacher interpretations of the task (Hannah Interview), and as this is also a factor considered in the literature in regards to writing prompts and assignments (Murphy & Ruth, 1993; Ruth & Murphy, 1988), it should be studied. Cardinal College uses the understanding (or misunderstanding) of the placement prompt as one indicator of both reading ability and familiarity with academic literacy (or its lack), but testing this assumption would be useful to the program's future work. Do different interpretations of the prompt indicate something about student's potential for struggle with course assignments? Does direct work on understanding assignments, which typically takes place in the writing courses, make this factor irrelevant for placement decisions?

Several teachers expressed concerns with the difficulty level of the task and the readings. One teacher felt the student's lack of interest in the topic and understandings of the readings, as well as the heavily excerpted nature of the one reading, made the task a poor choice, as students would not best display their writing abilities (Terence Interview). Yet, another teacher thought the journalistic reading was too easy, and should be replaced with the use of two fully academic texts (Odysseus Interview). Since it is possible that too difficult a task can make students seem weaker than they are or place students into the lower course unnecessarily while too easy a task can make students appear more prepared than they are, these are important aspects of the placement prompt to consider as well.

Value of the Study for Cardinal College

This study was valuable to Cardinal College because we had an emerging set of concerns that needed to be addressed through ongoing validation research. First, we operated on the assumption (and have made the argument to administrators) that we use course criteria to make placement decisions, and we needed to explore that assumption as it was important to have data to support the assertions of our interpretive argument and to address the concerns of our stakeholders. Second, several years ago (around 2008), a four-credit course in the program was eliminated to regulate credit hours, and faculty have continually felt concerned about the loss of that course and the potential for students to not fit as neatly into the remaining course options.

Validation work has provided the program with meaningful understandings of its courses, and an in-depth, qualitative view of its teachers and students. Additionally, validation research has helped the program start to investigate its placement assumptions. Through this study, Cardinal College has been able to consider the way it meets the needs of its particular students, including those students who used to fit into the discontinued course, and how it might use future placement reading and departmental work to adjust the courses to better meet students' needs. Thus, the current study has had a direct and potentially profound effect on the work of Cardinal College's writing program even as the study has revealed the need to reevaluate the courses and improve the placement reading process. Having the clear and detailed description of the placement and course determinations and the many student examples presented here will serve the program well as it decides how best to reconfigure the courses, and as it reestablishes placement reading around new courses boundaries.

Value of the Study for the Field

The description and discussion of Cardinal College that was produced through this study allowed the program to study itself—to check the effectiveness of its procedures and study the validity of its decisions. By looking closely at how the criteria under discussion by the teachers in the program related and aligned, yet did not align to those under discussion during placement, the program benefitted. Yet additionally, this work, as I see it, was valuable in three other ways, each larger than the direct benefit to Cardinal College specifically.

Those With Similar Contexts and Concerns

Since validation research requires that the interpretive argument that is sketched be specific to the particular site and context, this study describes in detail the College, its courses, its students, and its teachers, as well as the goals and vision of its writing program. I contend that, for those who share similar teaching and learning contexts or concerns, such a detailed description and discussion can be instructive or helpful to their own situations. Specifically, I can imagine three situations where others might share Cardinal College's context or concerns. First, Cardinal College is a small liberal arts college with a writing program that is concerned primarily with teaching writing to undergraduates. There are many other writing programs who share that context and who wish to be able to conduct scholarship—particularly assessment scholarship—that doesn't take away from their primary work of teaching but instead feeds it (Donahue & Moon, 2007; Gladstein & Regaignon, 2012). The study here provides a procedure and model for writing programs at these institutions to study their own practices and improve their teaching and learning goals—one step at a time—through ongoing validation.

Second, those writing programs that have integrated reading and writing courses can learn a great deal from the criteria developed in this study—particularly the way teachers understand students’ misreading, how they theorize students must conceive of the purpose or tasks of reading, how they weigh the importance of reading issues against writing ones, and how they imagine the impact reading and writing have on each other. Relatedly, since the first finding of this study showed that Cardinal College’s program, readers, and courses see students on a continuum of linked reading and writing ability, those already studying how students’ reading and writing development impact each other can work with and add to the developmental continuums produced from the integrated reading and writing criteria uncovered in this study.

Third, for writing programs who are worried about some type of disagreement or misalignment in the context of their site, this study provides a glimpse at how to use validation to study existing misalignments. In the case of Cardinal College, the misalignment was due to shifting course boundaries brought about by unplanned curricular change, but not every context has a lost course. The example provided by this study can help others make sense of their specific institutional histories or teaching differences which may have contributed to misalignment or disagreement in some form. Potential disagreement or misalignment comes about from many situations—teachers can disagree on assessment decisions, a program can disagree with the administration on some course of action or practice, a program can simply want to confirm course boundaries or alignments, or a program could want to study some specific aspect of its particular institutional history. The details offered here can help others think through their assumptions and work.

Curricular Change

The findings uncovered suggest something about the importance of the work of assessment during times of curricular change. The study of Cardinal College's specific situation, with its example of a program making placement decisions after the elimination of a course offering, is particularly instructive on the importance of validation work in times of curricular change generally. As assessments are only valid in the context of their intended use, and as we are responsible for checking that the assumptions we've made in the design and implementation of our assessments are aligned with reality, the process of curricular change requires substantial validation work. Yet the usefulness of validation in times of curricular change, particularly externally required change, is deeply contextual as well.

It is difficult for teachers and programs to know, when faced with such changes, what they must do to best meet the needs of their students. It is possible to have teaching knowledge and expertise that will find no outlet, even through a placement reading model like Smith's (1992/2009) expert-reader model which presumes to tap into teachers' expertise, when the courses themselves are not aligned with what teachers have come to know about their particular student population. Barritt, Stock, and Clark (1986) found that forced discussion of pre-existing criteria in assessment readings achieved agreement among readers, but these forced criteria made it hard for teachers to take themselves seriously as experts so they could engage with and make sense of what they knew about students and teaching. The same can be true in the opposite direction: a forced curricular structure can also make it hard for teachers to seriously engage with their expertise. In these moments, teacher's existing expertise and the key understandings contained in their

conceptual frameworks, built over many years, can seem more like a burden when they only help teacher's note pedagogically relevant issues that can find no clear curricular outlet.

It is in these moments that teachers and programs must fight to continue making decisions that improve teaching and learning, even when their choices are not ideal. Teachers must think about what work can best help them meet the needs of their students in the face of existing contextual constraints. They must consider which decisions will allow them to preserve their ability to do work they believe in. Validation work can help teachers develop contextualized understandings and interpretations, but it can also help them to continue engaging with their expertise in order to figure out how to find an outlet and expression for what they know. Through validation, teachers can identify what isn't working, gather support for their expert interpretations, and make changes.

Culture of Teaching-and-Learning-Based Writing Assessment

Finally, the study of specific sites like this one, with published results discussing the work toward validating the decisions being made, helps in the creation and promotion of a culture in composition where writing assessment is done with a focus on locally controlled, site-based assessments and their particular teaching and learning goals and contexts. Current writing assessment theory (CCCC, 2009; Huot, 1996/2009, 2002; Huot & O'Neill, 2009) implores us to make a case for, through our own examples, the kind of writing assessment that aids the real work of teaching and learning writing. Yet, the publication of such studies is comparatively rare in the literature. This study adds another example of such work to the literature of writing assessment. As it presents a published example of a locally-created writing assessment and the work of validating it, this study

reasserts that assessment, to be well done, must be deeply connected to the specific teaching goals and contexts it is assessing. Examples like this one can challenge the mythoi of standardized assessments, decontextualized writing, and one-size-fits-all teaching and learning, and they can make it possible for teachers and programs to maintain the integrity of their work, in the face of external pressure, and to gather evidence to continue fighting for teaching and learning based writing assessment.

References

- Adams, P. D. (1993). Basic writing reconsidered. *Journal of Basic Writing*, 12(1), 22-35.
Retrieved from <http://wac.colostate.edu/jbw/>
- Allen, M. (1995). Valuing differences: Portnet's first year. *Assessing Writing*, 2(1), 67-90. [http://dx.doi.org/10.1016/1075-2935\(95\)90005-5](http://dx.doi.org/10.1016/1075-2935(95)90005-5)
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.

- Barritt, L., Stock, P. L., & Clark, F. (1986). Researching practice: Evaluating assessment essays. *College Composition and Communication*, 37(3), 315-327.
<http://dx.doi.org/10.2307/358048>
- Bartholomae, D. (1979). Teaching basic writing: An alternative to basic skills. *Journal of Basic Writing*, 2(2), 85-109. Retrieved from <http://wac.colostate.edu/jbw/>
- Bartholomae, D. (1987). Writing on the margins: The concept of literacy in higher education. In T. Enos (Ed.), *A sourcebook for basic writing teachers* (pp. 66-83). New York, NY: Random House.
- Bartholomae, D. (1993). The tidy house: Basic writing in the American curriculum. *Journal of Basic Writing*, 12(1), 4-21. Retrieved from <http://wac.colostate.edu/jbw/>
- Bartholomae, D. (2005). Inventing the university. In D. Bartholomae (Author), *Writing on the margins* (pp. 60-85). Boston, MA: Bedford/St. Martin's. (Reprinted from *When a writer can't write: Studies in writer's block and other composing-process problems*, pp.134-165, by M. Rose, Ed., 1985, New York, NY: Guilford Press)
- Bartholomae, D., & Petrosky, A. R. (1986). *Facts, artifacts, and counterfacts: Theory and method for a reading and writing course*. Upper Montclair, NJ: Boynton/Cook.
- Bartholomae, D., & Petrosky, A. R. (2004). *Ways of reading: An anthology for writers*. New York, NY: Bedford/St. Martin's.
- Belfield, C. R., & Crosta, P. M. (2012). *Predicting success in college: The importance of placement tests and high school transcripts* (CCRC Working Paper No. 42). Retrieved from Community College Research Center, Teachers College,

- Columbia University website: <http://ccrc.tc.columbia.edu/publications/predicting-success-placement-tests-transcripts.html>.
- Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). *Assessing writing skill*. New York, NY: College Entrance Examination Board.
- Bunn, M. (2011). How to read like a writer. In C. Lowe, & P. Zemliansky (Eds.), *Writing spaces: Readings on writing* (Vol. 2, pp. 71-86). Anderson, SC: Parlor Press.
- Camp, R. (1993). Changing the model for the direct assessment of writing. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 45-78). Cresskill, NJ: Hampton Press.
- Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, 18(1), 65-81.
- Cherry, R. D., & Meyer, P. R. (1993). Reliability issues in holistic assessment. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 109-141). Cresskill, NJ: Hampton Press.
- Coles, W. E., Jr. (1988). *The plural I—and after*. Portsmouth, NH: Boynton/Cook.
- Conference on College Composition and Communication. (2009, March). *Writing assessment: A position statement*. Retrieved from <http://www.ncte.org/cccc/resources/positions/writingassessment>
- Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper, & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 3-31) Urbana, IL: National Council of Teachers of English.

- Cooper, P. L. (1984). *The assessment of writing ability: A review of research* (GREEB No. 82-15). Princeton, NJ: Educational Testing Service.
- Creswell, J. W. (2007). *Qualitative inquiry and research design* (2nd ed.). Thousand Oaks, CA: Sage.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1980). Validity on parole: How can we go straight? In W. B. Schrader (Ed.), *New Directions for Testing and Measurement: No. 5. Measuring Achievement: Progress Over a Decade, Proceedings of the 1979 ETS Invitational Conference* (pp. 99-108). San Francisco: Jossey-Bass.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer, & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L. J. (1989). Construct validity after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement theory and public policy* (pp. 147-171). Urbana: University of Illinois Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <http://dx.doi.org/10.1037/h0040957>
- Crowley, S. (1996). Response to Edward M. White's "The importance of placement and basic studies." *Journal of Basic Writing*, 15(1), 88-91. Retrieved from <http://wac.colostate.edu/jbw/>
- Davis, B. G., Scriven, M., & Thomas, S. (1981). *The evaluation of composition instruction*. Inverness, CA: Edgepress.

- Denzin, N. K., & Lincoln, Y. S., Eds. (2008). *Collecting and interpreting qualitative materials* (3rd ed.). Thousand Oaks, CA: Sage.
- Diederich, P. B., French, J. W., & Carlton, S. T. (1961). *Factors in judgments of writing ability*. (ETS Research Bulletin No. 61-15). <http://dx.doi.org/10.1002/j.2333-8504.1961.tb00286.x>
- Donahue, P., & Moon, G. F. (2007). *Local histories: Reading the archives of composition*. Pittsburgh, PA: University of Pittsburgh Press.
- Durst, R., Roemer, M., & Shultz, L. (1994). Portfolio negotiations: Acts in speech. In L. Black, D. A. Daiker, J. Sommers, & G. Stygall (Eds.), *New directions in portfolio assessment: Reflective practice, critical theory, and large-scale scoring* (pp. 286-300). Portsmouth, NH: Boynton-Heinemann.
- Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. New York, NY: Peter Lang.
- Elliot, N., Deess, P., Rudiny, A., & Joshi, K. (2012). Placement of students into first-year writing courses. *Research in the Teaching of English*, 46(3), 285-313.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251. <http://dx.doi.org/10.1037/0033-295x.87.3.215>
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Faigley, L., Cherry, R. D., Jolliffe, D. A., & Skinner, A. M. (1985). *Assessing writers' knowledge and processes of composing*. Norwood, NJ: Ablex.
- Fish, S. E. (1980). *Is there a text in this class? The authority of interpretive communities*. Cambridge, MA: Harvard University Press.

- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387. <http://dx.doi.org/10.2307/356600>
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37(1), 16-55. <http://dx.doi.org/10.2307/357381>
- Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy, and Practice* 10(1), 27-46. <http://dx.doi.org/10.1080/09695940301696>
- Gere, A. R. (1980). Written composition: Toward a theory of evaluation. *College English* 42(1), 44-58. <http://dx.doi.org/10.2307/376032>
- Gere, A. R., Aull, L., Green, T., & Porter, A. (2010). Assessing the validity of directed self-placement at a large university. *Assessing Writing*, 15(3), 154-176. <http://dx.doi.org/10.1016/j.asw.2010.08.003>
- Gere, A. R., Aull, L., Perales Escudero, M. D., Lancaster, Z., & Lei, E. V. (2013). Local assessment: Using genre analysis to validate directed self-placement. *College Composition and Communication*, 64(4), 605-633.
- Gladstein, J. M., & Regaignon, D. R. (2012). *Writing program administration at small liberal arts colleges*. Anderson, SC: Parlor Press.
- Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). *The measurement of writing ability*. New York, NY: College Entrance Examination Board. [http://dx.doi.org/10.1002/1520-6807\(196701\)4:1%3C93::aid-pits2310040126%3E3.0.co;2-y](http://dx.doi.org/10.1002/1520-6807(196701)4:1%3C93::aid-pits2310040126%3E3.0.co;2-y)

- Graff, G., Birkenstein, C., & Durst, R. (2012). *"They say/I say": The moves that matter in academic writing with readings* (2nd ed.). New York, NY: Norton.
- Greenberg, K. L. (1987). Research on basic writers: Theoretical and methodological issues. In T. Enos (Ed.), *A sourcebook for basic writing teachers* (pp. 187-207). New York, NY: Random House.
- Greenberg, K. L. (1992). Validity and reliability: Issues in the direct assessment of writing. *Writing Program Administration*, 16(1-2), 7-23. Retrieved from <http://www.wpacouncil.org/journalarchives>
- Greenberg, K. L. (1993). The politics of basic writing. *Journal of Basic Writing*, 12(1), 64-71. Retrieved from <http://wac.colostate.edu/jbw/>
- Guion, R. M. (1980). On trinitarian doctrines of validity. *Professional Psychology*, 11(3), 385-398. <http://dx.doi.org/10.1037/0735-7028.11.3.385>
- Hake, R. (1986). How do we judge what they write? In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 153-167). New York, NY: Longman.
- Harrington, S. (1998). New visions of authority in placement test rating. *WPA: Writing Program Administration*, 22(1-2), 53-84. Retrieved from <http://www.wpacouncil.org/journalarchives>
- Harrington, S. (2005). Learning to ride the waves: Making decisions about placement testing. *WPA: Writing Program Administration*, 28(3), 9-29. Retrieved from <http://www.wpacouncil.org/journalarchives>

- Hassel, H., & Giordano, J. B. (2011). First-year composition placement at open-admission, two-year campuses: Changing campus culture, institutional practice, and student success. *Open Words: Access and English Studies*, 5(2), 29-60.
- Haswell, R. (2005, March 1). Post-secondary entrance writing placement. Retrieved from: <http://comppile.org/profresources/placement.htm>
- Haswell, R., & Wyche-Smith, S. (1994). Adventuring into writing assessment. *College Composition and Communication*, 45(2), 220-236. <http://dx.doi.org/10.2307/359007>
- Hayes, J. R., & Flower, L. S. (1983). Uncovering cognitive processes in writing: An introduction to protocol analysis. In P. Mosenthal, L. Tamor, & S. A. Walmsley (Eds.), *Research on writing: Principles and methods* (pp. 207-220). New York, NY: Longman.
- Hayes, M. F., & Daiker, D. A. (1984). Using protocol analysis in evaluating responses to student writing. *Freshman English News*, 13(2), 1-10.
- Hester, V., O'Neill, P., Neal, M., Edgington, A., & Huot, B. (2007). In P. O'Neill (Ed.), *Blurring boundaries: Developing writers, researchers, and teachers* (pp.61-90). Cresskill, NJ: Hampton Press.
- Hoetker, J. (1982). Essay examination topics and students' writing. *College Composition and Communication*, 33(4), 377-392. <http://dx.doi.org/10.2307/357950>
- Huot, B. A. (1988). *The validity of holistic scoring: A comparison of the talk-aloud protocols of expert and novice holistic raters*. (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 8817872)

- Huot, B. (1990a). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication*, 41(2), 201-213.
<http://dx.doi.org/10.2307/358160>
- Huot, B. (1990b). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60(2), 237-263.
<http://dx.doi.org/10.3102/00346543060002237>
- Huot, B. A. (1993). The influence of holistic scoring procedures on reading and rating student essays. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 206-236). Cresskill, NJ: Hampton Press.
- Huot, B. A. (1994). A survey of college and university placement practices. *WPA: Writing Program Administrator*, 17(3), 49-67. Retrieved from <http://www.wpacouncil.org/journalarchives>
- Huot, B. (1996). The need for a theory of writing assessment. In L. Z. Bloom, D. A. Daiker, & E. M. White (Eds.), *Composition in the twenty-first century: Crisis and change* (pp.112-115). Carbondale: Southern Illinois University Press.
- Huot, B. (2002). *(Re)articulating writing assessment for teaching and learning*. Logan: Utah State University Press.
- Huot, B. (2007). Standing on his shoulders: Understanding William L. Smith's contributions to writing assessment. In P. O'Neill (Ed.), *Blurring boundaries: Developing writers, researchers, and teachers* (pp.15-30). Cresskill, NJ: Hampton Press.

- Huot, B. (2009). Toward a new theory of writing assessment. In B. Huot, & P. O'Neill (Eds.), *Assessing writing: A critical sourcebook* (pp. 160-175). Boston, MA: Bedford/St. Martin's. (Reprinted from *College Composition and Communication*, 47[4], 549-566. 1996)
- Huot, B., & Neal, M. (2006). Writing assessment: A techno-history. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 417-432). New York, NY: Guilford Press.
- Huot, B., & O'Neill, P. (2009). *Assessing writing: A critical sourcebook*. Boston, MA: Bedford/St. Martin's.
- Kane, M. T. (1982). A sampling model for validity. *Applied Psychological Measurement*, 6(2), 125-160. <http://dx.doi.org/10.1177/014662168200600201>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535. <http://dx.doi.org/10.1037//0033-2909.112.3.527>
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. <http://dx.doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-63). Westport, CT: Praeger-Greewood.
- Larson, R. L. (2000). Revision as self-assessment. In J. B. Smith, & K. B. Yancey (Eds.), *Self-Assessment and development in writing: A collaborative inquiry* (pp.97-103). Cresskill, NJ: Hampton Press.

- Leki, I. (1991). A new approach to advanced ESL placement testing. *WPA: Writing Program Administration*, 14(3), 53-68. Retrieved from <http://www.wpacouncil.org/journalarchives>
- Lewin, T. (2012, February 28). Colleges misassign many to remedial classes, studies find. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), 635-694. (Monograph Supplement 9).
<http://dx.doi.org/10.2466/pr0.1957.3.3.635>
- Lynne, P. (2004). *Coming to terms: A theory of writing assessment*. Logan: Utah State University Press.
- Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (Research Report 2009-2). Retrieved from College Board website: <http://professionals.collegeboard.com/gateway>.
- Maxwell, J. A. (1996). *Qualitative research design*. Thousand Oaks, CA: Sage.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30(10), 955-966. <http://dx.doi.org/10.1037/0003-066x.30.10.955>
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10(9), 9-20. <http://dx.doi.org/10.3102/0013189x010009009>
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer, & H. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum.

- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-12. <http://dx.doi.org/10.3102/0013189x018002005>
- Messick, S. (1989b). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Messick, S. (1995). The standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8. <http://dx.doi.org/10.1111/j.1745-3992.1995.tb00881.x>
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performative assessment. *Review of Educational Research*, 62(3), 229-258. <http://dx.doi.org/10.3102/00346543062003229>
- Moss, P. A. (1994). Can there be validity without reliability. *Educational Researcher*, 23(2), 5-12. <http://dx.doi.org/10.3102/0013189x023002005>
- Murphy, S., & Ruth, L. (1993). The field testing of writing prompts reconsidered. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 266-302). Cresskill, NJ: Hampton Press.
- Murphy, S., & Yancey, K. B. (2008). Construct and consequence: Validity in writing assessment. In C. Bazerman (Ed.), *Handbook of research on writing* (pp. 365-386). New York, NY: Lawrence Erlbaum.
- Myers, M. (1980). *A procedure for writing assessment and holistic scoring*. Urbana, IL: National Council of Teachers of English.

- Neal, M., & Huot, B. (2003). Responding to directed self-placement. In D. J. Royer, & R. Gilles (Eds.), *Directed self-placement: Principles and practices* (pp. 243-256). Cresskill, NJ: Hampton Press.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231-259. <http://dx.doi.org/10.1037/0033-295x.84.3.231>
- Odell, L. (1981). Defining and assessing competence in writing. In C. Cooper (Ed.), *The nature and measurement of competency in English* (pp. 95-138). Urbana, IL: National Council of Teachers of English.
- O'Neill, P. (2011). Reframing reliability for writing assessment. *Journal of Writing Assessment*, 4(1). Retrieved from <http://www.journalofwritingassessment.org/>
- O'Neill, P., Moore, C., & Huot, B. (2009). *A guide to college writing assessment*. Logan: Utah State University Press.
- Otte, G., & Mlynarczyk, W. (2010). *Basic writing*. West Lafayette, IN: Parlor Press and WAC Clearinghouse.
- Palmer, O. (1961). Sense or nonsense? The objective testing of English composition. *English Journal*, 50(5), 314-320. <http://dx.doi.org/10.2307/810351>
- Parkes, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26(4), 2-10. <http://dx.doi.org/10.1111/j.1745-3992.2007.00103.x>
- Pula, J. J., & Huot, B. A. (1993). A model of background influences on holistic raters. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 237-265). Cresskill, NJ: Hampton Press.

- Purves, A. C. (1995). Apologia not accepted. *College Composition and Communication*, 46(4), 549-550. <http://dx.doi.org/10.2307/358328>
- Royer, D. J., & Gilles, R. (1998). Directed self-placement: An attitude of orientation. *College Composition and Communication*, 50(1), 54-70. <http://dx.doi.org/10.2307/358352>
- Royer, D. J., & Gilles, R. (2003). *Directed self placement: Principles and practices*. Cresskill, NJ: Hampton Press.
- Ruth, L., & Murphy, S. (1988). *Designing writing tasks for the assessment of writing*. Norwood, NJ: Ablex.
- Sales, B. D., & Folkman, S. (2000). *Ethics in research with human participants*. Washington, DC: American Psychological Association.
- Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). Retrieved from Community College Research Center, Teachers College, Columbia University website: <http://ccrc.tc.columbia.edu/publications/high-stakes-placement-exams-predict.html>
- Shaughnessy, M. P. (1977). *Errors and expectations*. New York, NY: Oxford University Press.
- Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammon (Ed.), *Review of Research in Education* (Vol. 19, pp. 405-450). Washington, DC: American Educational Research Association.
- Shepard, L. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 1066-1101). Washington, DC: American Educational Research Association.

- Shor, I. (1997). Our apartheid: Writing instruction and inequality. *Journal of Basic Writing*, 16(1), 91-104. Retrieved from <http://wac.colostate.edu/jbw/>
- Smith, W. L. (2009). The importance of teacher knowledge in college composition placement testing. In B. Huot, & P. O'Neill, *Assessing writing: A critical sourcebook* (pp. 179-202). Boston, MA: Bedford/St. Martin's. (Reprinted from *Reading empirical research studies: The rhetoric of research*, pp. 289-316, by R. J. Hayes, Ed., 1992, Hillsdale, NJ: Lawrence Erlbaum)
- Smith, W. L. (1993). Assessing the reliability and adequacy of using holistic scoring of essays as a college composition placement program technique. In M. M. Williamson, & B. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 142-205). Cresskill, NJ: Hampton Press.
- Smith, W. L., Hull, G. A., Land, R. E., Jr., Moore, M. T., Ball, C., Dunham, D. E.,...Ruzich, C. W. (1985). Some effects of varying the structure of a topic on college students writing. *Written Communication*, 2(1), 73-89.
<http://dx.doi.org/10.1177/0741088385002001005>
- Snyder, T. D., & Dillow, S. A. (2012). *Digest of education statistics 2011* (NCES 2012-001). Retrieved from National Center for Education Statistics website:
<http://nces.ed.gov/pubs2012/2012001.pdf>
- Snyder, T. D., & Dillow, S. A. (2015). *Digest of education statistics 2013* (NCES 2015-011). Retrieved from National Center for Education Statistics website:
<http://nces.ed.gov/pubs2015/2015011.pdf>

- Sommers, N. (1993). I stand here writing. *College English*, 55(4), 420-428.
<http://dx.doi.org/10.2307/378651>
- Sommers, N., & Saltz, L. (2004). The novice as expert: Writing the freshman year. *College Composition and Communication*, 56(1), 124-149. <http://dx.doi.org/10.2307/4140684>
- Stock, P. L., & Robinson, J. L. (1987). Taking on testing: Teachers as tester-researchers. *English Education*, 19(2), 93-121.
- Suppe, P. (1977). *The structure of scientific theories*. Urbana: University of Illinois Press.
- Swarts, H., Flower, L., & Hayes, J. R. (1984). Designing protocol studies of the writing process. In R. Beach, & L. Bridwell (Eds.), *New directions in composition research* (pp. 53-71). New York, NY: Guilford
- Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology*, 30(1), 47-54. <http://dx.doi.org/10.1111/j.1744-6570.1977.tb02320.x>
- Thorndike, E. L. (1911). A scale for merit in English writing by young people. *Journal of Educational Psychology*, 2(7), 361-368. <http://dx.doi.org/10.1037/h0071519>
- Tierney, R. J., & Pearson, P. D. (1983). Toward a composing model of reading. *Language Arts*, 60(5), 568-580.
- Troyka, L. Q. (1987). Defining basic writing in context. In T. Enos (Ed.), *A sourcebook for basic writing teachers* (pp. 2-15). New York, NY: Random House.
- White, E. M. (1973). *Comparison and contrast: The 1973 California State University and Colleges Freshman English Equivalency Examination* (ERIC No. ED114825). Retrieved from ERIC, Institute of Education Services website: <http://www.eric.ed.gov/>

- White, E. M. (1986). Pitfalls in the testing of writing. In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 53-78). New York, NY: Longman.
- White, E. M. (1989). *Developing successful college writing programs*. San Francisco, CA: Jossey-Bass.
- White, E. M. (1993). Holistic scoring: Past triumphs, future challenges. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 79-108). Cresskill, NJ: Hampton Press.
- White, E. M. (1994). *Teaching and assessing writing: Recent advances in understanding, evaluating, and improving student performance* (2nd ed.). San Francisco, CA: Jossey-Bass.
- White, E. M. (1995a). An apologia for the timed impromptu essay test. *College Composition and Communication*, 46(1), 30-45. <http://dx.doi.org/10.2307/358868>
- White, E. M. (1995b). Apologia not accepted: Response. *College Composition and Communication*, 46(4), 550-551. <http://dx.doi.org/10.2307/358329>
- White, E. M. (1995c). The importance of placement and basic studies: Helping students succeed under the new elitism. *Journal of Basic Writing*, 14(2), 75-84. Retrieved from <http://wac.colostate.edu/jbw/>
- White, E. M. (2001). The opening of the modern era of writing assessment. *College English*, 63(3), 306-320. <http://dx.doi.org/10.2307/378995>

- Wiggins, G. (1994). The constant danger of sacrificing validity to reliability: Making writing assessment serve writers. *Assessing Writing*, 1(1), 129-140.
[http://dx.doi.org/10.1016/1075-2935\(94\)90008-6](http://dx.doi.org/10.1016/1075-2935(94)90008-6)
- Williamson, M. M. (1993). An introduction to holistic scoring: The social, historical, and theoretical context for writing assessment. In M. M. Williamson, & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 1-43). Cresskill, NJ: Hampton Press.
- Williamson, M. M., & Huot, B. A. (1993). *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. Cresskill, NJ: Hampton Press.
- Witte, S. P. (1987). Pre-text and composing. *College Composition and Communication*, 38(4), 397-425. <http://dx.doi.org/10.2307/357634>

Appendix A

Writing Placement Prompt

College Writing Writing Placement Exam

DIRECTIONS:

Read the attached passages and write an essay in response to what you have read. Your responses should be handwritten on one side of each page.

Your essay will be evaluated on your ability to:

- Provide a defensible reading of the assigned texts
- Write an essay that responds to the assigned question
- Make a thoughtful argument in writing
- Use quotations correctly
- Make connections between what you have read and other ways of seeing the same idea
- Control error in your writing

ESSAY QUESTION:

Attached you will find an excerpt from “Hidden Intellectualism” a short essay by Gerald Graff published in *They Say/ I say*, a book about academic writing by Graff and Cathy Birkenstein. You will also find an article from the *Pittsburgh Post-Gazette* about the recent “demotion” of Pluto from its status as a planet. Please read these materials carefully, making any marks or notes on the pages that you think will help you develop your reading and your essay.

Write an essay in which you explain your understanding of what Graff calls the “real intellectual world” and how it works. As you think about his idea, you may want to consider why he believes this world is “kept hidden from students” and what those tragic consequences might be.

Once you have worked out your reading of Graff’s argument, go on to apply it to an event that took place recently in the “real intellectual world”: the demotion of Pluto from a planet to a “dwarf planet”. How might Graff’s ideas better help you understand what the big deal is about “demoting” Pluto?

Here is an excerpt from Gerald Graff's "Hidden Intellectualism":

It was in these discussions with friends about toughness and sports I think, and in my reading of sports books and magazines, that I began to learn the rudiments of the intellectual life: how to make an argument, weigh different kinds of evidence, move between particulars and generalizations, summarize the views of others, and enter a conversation about ideas. It was in reading and arguing about sports and toughness that I experienced what it felt like to propose a generalization, restate and respond to a counterargument, and perform other intellectualizing operations, including composing the kind of sentences I am writing now,

Only much later did it dawn on me that the sports world was more compelling than school because it was more intellectual than school, not less. Sports after all was full of challenging arguments, debates, problems for analysis, and intricate statistics that you could care about, as school conspicuously was not....

For here is another thing that never dawned on me and is still kept hidden from students, with tragic results: that the real intellectual world, the one that existed in the big world beyond school, is organized very much like the world of team sports, with rival texts, rival interpretations and evaluations of texts, rival theories of why they should be read and taught, and elaborate team competitions in which "fans" of writers, intellectual systems, methodologies and -isms contend against each other. (300-1)

Post-gazette.com Science

**Then there were 8: Pluto isn't a planet
Astronomers' New Rules Change Status**

Friday, August 25, 2006

By Pete Zapadka, *Pittsburgh Post-Gazette*

Pluto arrived on the public's consciousness in 1930 when Clyde Tombaugh, an Illinois farm boy turned astronomer, discovered the frigid world after an extensive search for a much-sought "Planet X." Yesterday, however, Pluto's status as a planet came crashing down to earth. In its annual convention in Prague, Czech Republic, the International Astronomical Union revoked Pluto's planetary status under the new guidelines meant to bring the definition of "planet" in line with our expanding technology. Pluto doesn't make the grade under the new rule: "a celestial body that is in orbit around the sun, has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a ... nearly round shape, and has cleared the neighborhood around its orbit."

Pluto is disqualified because its oblong orbit overlaps with Neptune's. It will now be reclassified in a new category of "dwarf planets," similar to what long have been termed "minor planets." The definition also lays out a third class of lesser objects that orbit the sun—"small solar system bodies," a term that will apply to numerous asteroids, comets and other natural satellites.

Astronomers feared that, by declaring Pluto and similar worlds such as its moon Charon, the largest asteroid Ceres and 2003 UB313 as planets, that the solar system in years to come might be overrun with planets as new discoveries are made. Ceres and 2003UB313 qualify as dwarfs; Charon may not receive any special designation.

"I think it's probably a good thing," Dr. Bruce Hapke said of Pluto's demotion.

Dr. Hapke, professor emeritus of geology and planetary science at the University of Pittsburgh, studies the surfaces of planets and satellites by analyses of data taken from the Earth and by spacecraft. He was a member of the Mariner 10 and Viking imaging science teams.

"If we continue to call Pluto a planet, there's probably a lot of bodies the size of Pluto out beyond the orbit of Neptune that we haven't discovered yet," Dr. Hapke said. "If you call Pluto a planet you'd have to call these things a planet, and we'd end up with a solar system with a hundred planets."

While the IAU's decision is deemed widely by the astronomical community as scientifically sound, it may not sit well with the general public, which has an affinity for Pluto and will have to rethink a solar system consisting of eight planets instead of nine.

It's just a question of tradition," Hapke said. "People are always upset when tradition gets changed. But you have to change as you learn more about something."

Appendix B

Email to Placement Readers

Dear Placement Reader,

As you may already know, I am conducting a study on our writing placement process here at the College investigating the connection between placement reading criteria and course criteria. I would like to invite you to participate in the research study since you are involved with writing placement as a placement reader. Attached you will find an informed consent form that details information such as the focus of the study, the requirements of participation, and the potential risks and benefits. This information will help you as you decide whether or not you would like to participate.

Once you've had some time to read this form, I would like to set up a meeting where we can discuss any questions you may have about the study. Please let me know when it is convenient for you to schedule this meeting.

Thank you, and I look forward to speaking with you.

Jessica Ganni

Appendix C

Informed Consent Form: Placement Readers

Working Title: Validating Placement for Teaching and Learning

You are invited to participate in this research study. The following information is provided in order to help you make an informed decision whether or not to participate. If you have any questions, please do not hesitate to ask. You are eligible to participate because you are a teacher who takes part in the reading of incoming students' placement essays at [Cardinal College].

The purpose of this study is to collect validation evidence for the placement process at [Cardinal College] by investigating the connection between the criteria that determine course placement and the criteria discussed by course teachers. Your participation in this study will involve allowing me (1) to conduct and audio record an initial interview (10-20 minutes) about your teaching background and experience; (2) to watch and audio record you as you read and think aloud about student placement essays during placement sessions; and (3) to re-interview you (no longer than 60 minutes) about criteria that emerge during the analysis of earlier sessions, if necessary. Additionally, your participation will allow me (4) to transcribe audio recordings of you into a written record; (5) to analyze the transcripts for recurring themes in the criteria used to determine course placement; and (6) to extract and use quotations from your responses to illustrate and document placement criteria. Your identity will be kept private, and a pseudonym of your choice will be used instead of your name on all materials. A master file linking your name to your pseudonym will be kept in a password protected file and destroyed upon completion of the study. You will be given the transcripts of all audio recordings prior to analysis for you to review so that transcripts accurately reflect your intended meaning. All audio recordings will be destroyed once they are fully transcribed and you have approved them.

There are no known risks associated with this research, but thinking aloud about essays may increase reading time slightly. You may find the interview and the experience of thinking aloud as you read student work enjoyable and helpful to your work teaching and determining placement for writing courses. The information gained from this study may help us to better understand, adjust, and review the placement process at [Cardinal College], and it may add to what is known about the value of local writing assessments.

Your participation in this study is voluntary. You are free to decide not to participate in this study or to withdraw at any time without adversely affecting your relationship with the investigators, the writing program, or [Cardinal College]. If you choose to participate, you may withdraw at any time by notifying either the Researcher or Faculty Sponsor. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all information will be held in strict confidence and will have no bearing on your standing or relationship with the College. The information obtained in

this study may be published in scholarly works or presented at scholarly meetings but your identity will be kept strictly confidential.

If you are willing to participate in this study, please sign a copy of the form and return it to me. Keep an extra unsigned copy with you. If you choose not to participate, return the unsigned copies to me.

Researcher:

Jessica A. Ganni
PhD Candidate in English
Indiana University of Pennsylvania
2324 Saranac Ave.
Pittsburgh, PA 15216
412-389-3458
jessica.ganni@[cardinal].edu

Faculty Sponsor:

Dr. Michael M. Williamson
Leonard Hall, Room 110B
Indiana University of Pennsylvania
Indiana, PA 15701
724-357-2671
mmwimson@iup.edu

This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (phone: 724-357-7730)

VOLUNTARY CONSENT FORM:

I have read and understand the information on the form, and I consent to volunteer to be a subject in this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time. I have received an unsigned copy of this informed consent form to keep in my possession.

Name (please print): _____

Signature: _____

Date: _____ Phone where you can be reached: _____

Best days and times to reach you: _____

I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, have answered any questions that have been raised, and have witnessed the above signature.

Date: _____ **Investigator's signature:** _____

Appendix D

Informed Consent Form: Placement Students

Working Title: Validating Placement for Teaching and Learning

You are invited to participate in this research study. The following information is provided in order to help you make an informed decision whether or not to participate. If you have any questions, please do not hesitate to ask. You are eligible to participate because you are taking the writing placement test at [Cardinal College].

The purpose of this research is to study the effectiveness of [Cardinal College's] writing placement testing process by checking that the kinds of work required in the writing courses are related to the kinds of work being examined during placement. Your participation in this study will involve allowing me (1) to watch and record placement readers read and think aloud about your placement essay; (2) to copy and read your placement essay myself; and (3) to use quotations from your essay in the study, if necessary, as examples of the kinds of work placement readers discuss during placement testing. Your identity will be kept private and a code number will be used on all materials related to your essay instead of your name. A master file linking your name to that code number will be kept in a password protected file, and the file will be destroyed once the study has ended.

There are no known risks associated directly with this research since placement testing is already a natural part of attending the College and the study will not interfere in the natural reading process. Further, the focus of the study is not on your work individually but on the kinds of writing under discussion by teacher-readers at placement and in writing courses.

The information gained from this study may not benefit you directly, but the information may help us to better understand, adjust, and review the placement testing process for writing courses at [Cardinal College], and it may add to what is known about writing assessment.

Your participation in this study is voluntary. You are free to decide not to participate in this study or to withdraw at any time without adversely affecting your relationship with the investigators, the writing program, or [Cardinal College]. If you choose to participate, you may withdraw at any time by notifying either the Researcher or Faculty Sponsor. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all information will be held in strict confidence and will have no bearing on your academic standing or services you receive from the College. The information obtained in this study may be published in scholarly works or presented at scholarly meetings, but your identity will be kept strictly confidential.

If you are willing to participate in this study, please sign the statement below and deposit the form in the box by the door. Take the extra unsigned copy with you. If you choose not to participate, deposit the unsigned copies in the box by the door.

Researcher:

Jessica A. Ganni
PhD Candidate in English
Indiana University of Pennsylvania
2324 Saranac Ave.
Pittsburgh, PA 15216
412-389-3458
jessica.ganni@[cardinal].edu

Faculty Sponsor:

Dr. Michael M. Williamson
Leonard Hall, Room 110B
Indiana University of Pennsylvania
Indiana, PA 15701
724-357-2671
mmwimson@iup.edu

This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (phone: 724-357-7730)

VOLUNTARY CONSENT FORM:

I have read and understand the information on the form, and I consent to volunteer to be a subject in this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time. I have received an unsigned copy of this informed consent form to keep in my possession.

Name (*please print*): _____

Signature: _____

Date: _____ **Phone where you can be reached:** _____

Best days and times to reach you: _____

I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, have answered any questions that have been raised, and have witnessed the above signature.

Date: _____ **Investigator's signature:** _____

Appendix E

Placement Reader Initial Interview Questions

- Please state the pseudonym you have chosen for the purposes of this research and your age.
- What is your educational background? Where did you study (undergraduate, graduate, etc.) and what degrees do you hold? What specifically did you study? Did you specialize in anything in particular inside your field?
- How long have you taught at the College? Where, if anywhere, did you teach before coming here?
- What first year writing course do you teach here at the college? What other courses do you teach?
- Describe your understanding of the goals, work of, and/or purpose of the writing program and the first-year courses overall.
- Describe the kinds of writing and reading work common to the course of your expertise (either ENGL1010 or ENGL1011) and what students should know or be able to do in this course. Specifically, what expectations do you have for students as they enter? What work does the course teach? What expectations do you have for students as they finish the course?
- How do you understand the placement prompt? What is it asking students to do? How does it relate to the course work of your course of expertise? Where do students in your course of expertise struggle or excel with the prompt?
- How do you approach placement reading? What are you reading essays for? What seems to you to be recurring aspects of interest as you read and place student work? What, in your mind currently, are some of the aspects that signal a student belongs in your course of expertise?

Appendix F

Think-Aloud Directions and Practice for Placement Readers

In these placement reading sessions, I am interested in what you say to yourself as you read these essays and decide if the students belong in your course or not. Therefore, I am asking you to think aloud as you read and make a decision. When I say think aloud, I mean I want you to speak out loud everything that goes through your mind. If you are silent for any length of time, I will remind you to keep thinking out loud. As you start each essay, begin by calling out the number of the essay. Do you understand what I want you to do?

Before you begin, I want you to practice thinking aloud. First, practice the work of thinking aloud as you answer a few questions: How many windows do you have in your house/apartment? How many words can you think of that rhyme with break?

Next, practice thinking aloud as you read and make a placement decision about this sample placement essay from an earlier placement session.

In a part of the essay “Hidden Intellectualism” in *They Say/I Say*, the author-Gerald Graff-initially elucidates the notion of “real intellectual world” in the context of sports world. He also argues for the necessary extension of the notion to schools and beyond schools. Applying Graff’s idea of the “real intellectual world” to the event of the demotion of Pluto to a “dwarf planet” is an example for the applicability of this definition.

First of all, Graff contends that the intellectual life should not be different between sports world and school. The real intellectual life is an integration of making an argument, critically evaluating different types of evidence, being flexible between particulars and generalizations, condensing the views of others, and participating a conversation about ideas. However, as a matter of fact, Graff has found that sports world is “more intellectual than school” (G. Graff). Sports world provokes people much more than school does since sports is “full of challenging arguments, debates, problems for analysis, and intricate statistics” (G. Graff). Moreover, how the “real intellectual world” works remains hidden from student, causing bad consequences to the way how students interact with the information they learn at school and out of school. The “real intellectual world” exists in any circumstances with opposing and arguable matters, which require students to have sufficient skills of reading, interpreting, evaluating, and contending. If they are not given opportunities to learn about the “real intellectual world” at school, they would have difficulties on competing to survive or develop when they are out of school, as well as evaluate and absorb the large amount of information feeding the public every day.

Graff's statement about the existence of contentions around an issue in daily life has proven in the event of the demotion of Pluto. Although the claim of demoting Pluto to a "dwarf planet" is based upon "scientifically sound" standards, the issue has not received all agreements from everybody who is interested in the event. In 2006, after more than seventy years of the discovery of Pluto, its status was revoked from a planet to a "dwarf planet" according to the new rules of defining planets by the International Astronomical Union (P. Zapadka, 2006). The demotion of Pluto was supported by many professionals like Dr. Hapke. Still, he was in favor of the new reasonable rules on the definition of planets, he still concerned about the public attitudes towards the change of tradition. Dr. Hapke recognized and analyzed the event upon two different backgrounds, professional and social, so that he was able to acknowledge the original reasons for the supporting and opposing sides of the event. On the other hand, another professional, Paul Hertz, did not show any judgment on the demotion of Pluto, not because he did not care but because the piece of information did not affect on his current work. Holding different concern, astronomer Reiland worried about the changes in teaching about planets even though he totally agreed with the new IAU's rule. It is apparent that even among people whose profession is same, different opinions do exist since they analyze the information according to their particular concerns.

In conclusion, seeing an issue in real life requires people to have sufficient intellectual skills. Such skills should be taught as soon as beginning of school, thereby ensuring students abilities to decide on any matters in life.

Appendix G

Email to Course Teachers

Dear Course Teacher,

As you may already know, I am conducting a study on our writing placement process here at the College investigating the connection between placement reading criteria and course criteria. I would like to invite you to participate in the research study since you are teaching either or both ENGL1010 or ENGL1011 at the College. Attached you will find an informed consent form that details information such as the focus of the study, the requirements of participation, and the potential risks and benefits. This information will help you as you decide whether or not you would like to participate.

Once you've had some time to read this form, I would like to set up a meeting where we can discuss any questions you may have about the study. Please let me know when it is convenient for you to schedule this meeting.

Thank you, and I look forward to speaking with you.

Jessica Ganni

Appendix H

Informed Consent Form: Course Teachers

Working Title: Validating Placement for Teaching and Learning

You are invited to participate in this research study. The following information is provided in order to help you make an informed decision whether or not to participate. If you have any questions, please do not hesitate to ask. You are eligible to participate because you are a teacher who teaches either ENGL1010 or ENGL1011 at [Cardinal College].

The purpose of this study is to collect validation evidence for the placement process at [Cardinal College] by investigating the connection between the criteria that determine course placement and the criteria discussed by course teachers. Your participation in this study will involve allowing me (1) to come to your class to ask for student volunteers; (2) to conduct and audio record an initial interview (30-40 minutes) about your teaching, background, and course design; (3) to watch and audio record you as you read and think aloud about students' diagnostic essays and one additional piece of their writing; and (4) to re-interview you (no longer than 60 minutes) about criteria that emerge during the analysis of think-aloud sessions, if necessary. Additionally, your participation will allow me (5) to transcribe audio recordings of you into a written record; (6) to analyze the transcripts for recurring themes in the criteria discussed during the teaching of ENGL1010 or ENGL1011; and (7) to extract and use quotations from your responses to illustrate and document course criteria. Also, you will (8) complete a survey twice during the semester on how well you feel the students in your class are placed overall. Your identity will be kept private and a pseudonym of your choice will be used instead of your name on all materials. A master file linking your name to your pseudonym will be kept in a password protected file and destroyed upon completion of the study. You will be given the transcripts of all audio recordings prior to analysis for you to review so that transcripts accurately reflect your intended meaning. All audio recordings will be destroyed once they are fully transcribed and you have approved them.

You may find the interview and the experience of thinking aloud as you read student work enjoyable and helpful to your teaching of writing courses. The information gained from this study may help us to better understand, adjust, and review the placement process at [Cardinal College], and it may add to what is known about the value of local writing assessments.

Your participation in this study is voluntary. You are free to decide not to participate in this study or to withdraw at any time without adversely affecting your relationship with the investigators, the writing program, or the College. If you choose to participate, you may withdraw at any time by notifying either the Researcher or Faculty Sponsor. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all information will be held in strict confidence and will have no

bearing on your standing or relationship with the College. The information obtained in this study may be published in scholarly works or presented at scholarly meetings but your identity will be kept strictly confidential.

If you are willing to participate in this study, please sign a copy of the form and return it to me. Keep an extra unsigned copy with you. If you choose not to participate, return the unsigned copies to me.

Researcher:

Jessica A. Ganni
PhD Candidate in English
Indiana University of Pennsylvania
2324 Saranac Ave.
Pittsburgh, PA 15216
412-389-3458
jessica.ganni@[cardinal].edu

Faculty Sponsor:

Dr. Michael M. Williamson
Leonard Hall, Room 110B
Indiana University of Pennsylvania
Indiana, PA 15701
724-357-2671
mmwimson@iup.edu

This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (phone: 724-357-7730)

VOLUNTARY CONSENT FORM:

I have read and understand the information on the form, and I consent to volunteer to be a subject in this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time. I have received an unsigned copy of this informed consent form to keep in my possession.

Name (*please print*): _____

Signature: _____

Date: _____ **Phone where you can be reached:** _____

Best days and times to reach you: _____

I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, have answered any questions that have been raised, and have witnessed the above signature.

Date: _____ **Investigator's signature:** _____

Appendix I

Informed Consent Form: Students

Working Title: Validating Placement for Teaching and Learning

You are invited to participate in this research study. The following information is provided in order to help you make an informed decision whether or not to participate. If you have any questions, please do not hesitate to ask. You are eligible to participate because you are a student enrolled in either ENGL1010 or ENGL1011 at [Cardinal College].

The purpose of this research is to study the effectiveness of [Cardinal College's] writing placement testing process by checking that the kinds of work required in the writing courses are related to the kinds of work being examined during placement. Your participation in this study will involve allowing me (1) to watch and record your teacher as he or she reads and thinks aloud about your first-day's essay and the draft of one formal paper in the course (written around week four); (2) to access your writing placement essay from the placement testing office and your placement reader records from the director of college writing to review them myself, and if necessary, have them reassessed by readers; and (3) to obtain and read your final course portfolio. Additionally, your participation in this study will allow me (4) to obtain your teacher's assessment of your final portfolio and your final course outcome and grade once the semester is over. Finally, your participation will allow me (5) to copy your writing and (6) to use quotations from your writing in the study, if necessary, as examples of the kinds of work course teachers discuss during their courses or the kinds of work done at placement. Your identity will be kept private and a code number will be used on all materials related to your work instead of your name. A master file linking your name to that code number will be kept in a password protected file, and the file will be destroyed once the study has ended. At no time will your teacher know if you have chosen to participate or not.

There are no known risks associated directly with this research since your teacher's reading of your work is already a natural part of the course and the study will not interfere in your teacher's natural reading process. Further, the focus of the study is not on your work individually but on the kinds of writing under discussion by teachers at placement and in writing courses.

The information gained from this study may not benefit you directly, but the information may help us to better understand, adjust, and review the placement testing process for writing courses at [Cardinal College], and it may add to what is known about writing assessment.

Your participation in this study is voluntary. You are free to decide not to participate in this study or to withdraw at any time without adversely affecting your relationship with the investigators, your teacher, or [Cardinal College]. If you choose to participate, you may withdraw at any time by notifying either the Researcher or Faculty Sponsor. Upon

your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all information will be held in strict confidence and will have no bearing on your academic standing or services you receive from the College. The information obtained in this study may be published in scholarly works or presented at scholarly meetings, but your identity will be kept strictly confidential.

If you are willing to participate in this study, please sign the statement below and deposit the form in the box by the door. Take the extra unsigned copy with you. If you choose not to participate, deposit the unsigned copies in the box by the door.

Researcher:

Jessica A. Ganni
PhD Candidate in English
Indiana University of Pennsylvania
2324 Saranac Ave.
Pittsburgh, PA 15216
412-389-3458
jessica.ganni@[cardinal].edu

Faculty Sponsor:

Dr. Michael M. Williamson
Leonard Hall, Room 110B
Indiana University of Pennsylvania
Indiana, PA 15701
724-357-2671
mmwimson@iup.edu

This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (phone: 724-357-7730)

VOLUNTARY CONSENT FORM:

I have read and understand the information on the form, and I consent to volunteer to be a subject in this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time. I have received an unsigned copy of this informed consent form to keep in my possession.

Name (*please print*): _____

Signature: _____

Date: _____ **Phone where you can be reached:** _____

Best days and times to reach you: _____

I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, have answered any questions that have been raised, and have witnessed the above signature.

Date: _____ **Investigator's signature:** _____

Appendix J

Course Teacher Initial Interview Questions

- Please state the pseudonym you have chosen for the purposes of this research and your age.
- What is your educational background? Where did you study (undergraduate, graduate, etc.) and what degrees do you hold? What specifically did you study? Did you specialize in anything in particular inside your field?
- How long have you taught at the College? Where, if anywhere, did you teach before coming here?
- What first year writing course do you teach here at the college? What other courses do you teach?
- Describe your understanding of the goals, work of, and/or purpose of the writing program and the first-year courses overall.
- Describe the kinds of writing and reading work common to the course of your expertise (either ENGL1010 or ENGL1011) and what students should know or be able to do in this course. Specifically, what expectations do you have for students as they enter? What work does the course teach? What expectations do you have for students as they finish the course?
- What is your familiarity with and how do you understand the placement prompt? Looking at it specifically, how do you understand what is it asking students to do? How does it relate to the course work of your course of expertise? Where would students in your course of expertise struggle or excel with this prompt? How does this prompt relate to the diagnostic essay prompt you give on the first day of your course to check placement?
- Describe your particular current approach to the first-year course you teach (either ENGL1010 or ENGL1011). How have you designed the course? What materials (texts, films, etc.) are you using? What are your assignments asking students to do or learn in terms of reading, writing, and thinking? How have you sequenced the course? As you describe, please detail why have you made these decisions.

Appendix K

Think-Aloud Directions and Practice for Course Teachers

In these reading sessions, I am interested in what you say to yourself as you read student diagnostics and decide if the students belong in your course or not OR as you read student work and respond to it. Therefore, I am asking you to think aloud as you read your students work. When I say think aloud, I mean I want you to speak out loud everything that goes through your mind. If you are silent for any length of time, I will remind you to keep thinking out loud. As you start each piece of student work, begin by calling out the number on the piece. Do you understand what I want you to do?

Before you begin, I want you to practice thinking aloud. First, practice the work of thinking aloud as you answer a few questions: How many windows do you have in your house/apartment? How many words can you think of that rhyme with break?

Next, practice thinking aloud as you read and respond to this sample essay written during a placement session.

In a part of the essay “Hidden Intellectualism” in *They Say/I Say*, the author-Gerald Graff-initially elucidates the notion of “real intellectual world” in the context of sports world. He also argues for the necessary extension of the notion to schools and beyond schools. Applying Graff’s idea of the “real intellectual world” to the event of the demotion of Pluto to a “dwarf planet” is an example for the applicability of this definition.

First of all, Graff contends that the intellectual life should not be different between sports world and school. The real intellectual life is an integration of making an argument, critically evaluating different types of evidence, being flexible between particulars and generalizations, condensing the views of others, and participating a conversation about ideas. However, as a matter of fact, Graff has found that sports world is “more intellectual than school” (G. Graff). Sports world provokes people much more than school does since sports is “full of challenging arguments, debates, problems for analysis, and intricate statistics” (G. Graff). Moreover, how the “real intellectual world” works remains hidden from student, causing bad consequences to the way how students interact with the information they learn at school and out of school. The “real intellectual world” exists in any circumstances with opposing and arguable matters, which require students to have sufficient skills of reading, interpreting, evaluating, and contending. If they are not given opportunities to learn about the “real intellectual world” at school, they would have difficulties on competing to survive or develop when they are out of school, as well as evaluate and absorb the large amount of information feeding the public every day.

Graff's statement about the existence of contentions around an issue in daily life has proven in the event of the demotion of Pluto. Although the claim of demoting Pluto to a "dwarf planet" is based upon "scientifically sound" standards, the issue has not received all agreements from everybody who is interested in the event. In 2006, after more than seventy years of the discovery of Pluto, its status was revoked from a planet to a "dwarf planet" according to the new rules of defining planets by the International Astronomical Union (P. Zapadka, 2006). The demotion of Pluto was supported by many professionals like Dr. Hapke. Still, he was in favor of the new reasonable rules on the definition of planets, he still concerned about the public attitudes towards the change of tradition. Dr. Hapke recognized and analyzed the event upon two different backgrounds, professional and social, so that he was able to acknowledge the original reasons for the supporting and opposing sides of the event. On the other hand, another professional, Paul Hertz, did not show any judgment on the demotion of Pluto, not because he did not care but because the piece of information did not affect on his current work. Holding different concern, astronomer Reiland worried about the changes in teaching about planets even though he totally agreed with the new IAU's rule. It is apparent that even among people whose profession is same, different opinions do exist since they analyze the information according to their particular concerns.

In conclusion, seeing an issue in real life requires people to have sufficient intellectual skills. Such skills should be taught as soon as beginning of school, thereby ensuring students abilities to decide on any matters in life.

Appendix L

Survey of Placement Adequacy

After you have read the work submitted by your students during class, please complete the following survey. Note: Please fill out this survey for each section you are teaching of ENGL1010 or ENGL1011.

Below you will find five columns. In each column, indicate the number of students in one particular section of your first-year writing course who you feel fit the description of students in that column. At the bottom, indicate the total number of students registered in the course.

Should Have Been Placed in Lower Course	Slightly Below	Prototypic of the Course	Slightly Above	Should Have Been Placed in Higher Course
---	----------------	-----------------------------	----------------	--

Total Number of Students in This Section of the Course _____

Appendix M

EOTR Course Results

Teacher's Name: _____ Course/Section: _____ No. of Students: _____

Please complete this form for every section of ENGL1010 or ENGL1011 that you teach. In the first column, please record the name of every student in the class. Then, as the course instructor, please fill out the second column with observations on each student's portfolio. Please make sure your comments are detailed enough to reveal your judgment and the reading/writing criteria that inform it. Write "No Portfolio Submitted" for any student who did not submit a final portfolio of work as required.

Since you will bring the folders of any ENGL1010 students who you feel can skip to ENGL1012 directly and any students who you feel may be failing to the EOTR meeting for committee review, you must fill out columns three and four for these students only by noting the concerns and comments of any reviewers and the final decision made regarding the student once each student's portfolio has been reviewed.

Finally, once you have had a chance to calculate and determine it, please use the last column to record the final course grade for every student in the class.

Student Name	Teacher's Comments on Portfolio	EOTR Reviewer's Comments on Portfolio (if applicable)	Decision	Final Course Grade

Appendix N

Summary of All Participating Students' Demographic Information

Student Code	Male	Female	ESL	HS-Dual Enrolled
1010-01-01	X		X	
1010-01-02	X		X	
1010-01-03	X		X	
1010-01-04		X		
1010-01-05	X		X	
1010-01-06		X		
1010-01-07		X		
1010-01-08	X			
1010-01-09	X		X	
1010-01-10	X			
1010-01-11	X			
1010-02-01		X		
1010-02-02		X		
1010-02-03	X			
1010-04-01	X		X	
1010-04-02	X		X	
1010-04-03	X		X	
1010-04-04	X		X	
1010-04-05		X		
1010-04-06	X			
1010-04-07	X		X	
1010-04-08	X			
1010-04-09	X		X	
1010-04-10		X		
1010-04-11		X		
1010-04-12	X			
1010-04-13		X		
1010-05-01		X		
1010-05-02	X			
1010-05-03		X		
1010-05-04	X			
1010-05-05		X		
1010-05-06	X			
1010-05-07		X		
1010-05-08		X		
1010-05-09		X		
1010-06-01		X	X	
1010-06-02		X		
1010-06-03	X			

Student Code	Male	Female	ESL	HS-Dual Enrolled
1010-06-04		X		
1010-06-05	X		X	
1010-06-06		X		
1010-06-07	X			
1010-06-08		X		
1010-06-09		X		
1010-06-10		X		
1010-06-11	X			
1010-06-12		X		
1010-06-13		X		
1010-06-14		X		
1010-07-01		X		
1010-07-02		X		
1010-07-03		X		
1010-07-04		X		
1010-07-05		X		
1010-07-06		X	X	
1010-07-07		X		
1010-07-08		X		
1010-07-09		X		
1011-00-01	X			
1011-02-01	X		X	
1011-02-02		X		X
1011-02-03		X		
1011-02-04	X			X
1011-02-05		X		
1011-02-06		X		
1011-02-07	X			
1011-02-08	X			
1011-03-01		X		
1011-03-02	X			
1011-03-03	X			
1011-03-04	X			
1011-03-05	X			
1011-03-06		X		
1011-03-07	X			
1011-03-08	X			
1011-03-09	X			
1011-03-10		X		
1011-03-11	X		X	
1011-03-12	X			
1011-03-13	X			
1011-05-01	X			

Student Code	Male	Female	ESL	HS-Dual Enrolled
1011-05-02	X			
1011-05-03	X			
1011-05-04		X		
1011-05-05	X			
1011-05-06		X		
1011-05-07	X		X	
1011-05-08		X		
1011-05-09		X		
1011-07-01	X			
1011-07-02	X			
1011-07-03		X		
1011-07-04	X			
1011-08-01	X			
1011-08-02		X		
1011-08-03		X		
1011-08-04		X	X	
1011-08-05		X	X	
1011-08-06	X			
1011-08-07		X		
1011-08-08		X		
1011-08-09		X		
1011-08-10		X	X	
1011-09-01	X			
1011-09-02	X		X	
1011-09-03	X			
1011-09-04		X		
1011-09-05		X		
1011-09-06		X		
1011-09-07		X		
1011-09-08		X		
1011-09-09	X			
1011-09-10		X		
1011-09-11	X			
1011-09-12	X			
1011-09-13	X			
1011-09-14		X		
1011-09-15		X		
TOTAL /119	57	62	21	2

Appendix O

Summary of Decisions on Participating Students in Original Placement Sessions

Student Code	Course 10	Course 11	Course 11A	Tweener
1010-01-06	X			
1010-01-11	X		X	
1010-02-01	X			
1010-04-08	X		X	
1010-06-02	X			
1010-06-05	X			
1010-06-10	X			
1010-06-11	X			
1010-07-01	X			
1010-07-04	X			
1010-07-06	X			
1010-07-09	X			
1011-00-01		X		
1011-02-03		X	X	
1011-03-02		X	X	
1011-03-03		X		
1011-03-06		X	X	X
1011-03-13		X		
1011-05-04		X	X	
1011-05-06		X		
1011-09-01		X	X	
1011-09-02		X		
1011-09-03		X		
1011-09-06		X	X	
1011-09-08		X		
1011-09-11		X		
TOTAL /26	12	14	8	1

Appendix P

Summary of Decisions on Participating Students in Redo Placement Sessions

Student Code	Course 10	Course 11	Course 11A	Tweener	Different Decision
1010-01-01	X				
1010-01-02	X				
1010-01-03	X				
1010-01-04	X				
1010-01-05	X				
1010-01-06	X				
1010-01-07	X		X		
1010-01-08		X			X
1010-01-10	X		X		
1010-01-11		X	X		X
1010-02-01	X				
1010-02-02	X		X		
1010-02-03		X			X
1010-04-01	X				
1010-04-03	X				
1010-04-04	X				
1010-04-05	X		X		
1010-04-06	X				
1010-04-08	X				
1010-04-09	X				
1010-04-10		X			X
1010-04-11	X		X	X	
1010-04-12	X				
1010-04-13	X		X	X	
1010-05-01	X				
1010-05-02	X		X		
1010-05-03		X	X	X	X
1010-05-04	X				
1010-05-05	X				
1010-05-06		X			X
1010-05-07	X				
1010-05-08	X		X		
1010-05-09		X	X		X
1010-06-01	X		X		
1010-06-02	X				
1010-06-03		X			X
1010-06-04	X				
1010-06-05	X				
1010-06-06	X				
1010-06-07	X		X		

Student Code	Course 10	Course 11	Course 11A	Tweener	Different Decision
1010-06-08		X	X		X
1010-06-09		X			X
1010-06-10	X				
1010-06-11	X				
1010-06-12	X		X		
1010-06-13	X		X		
1010-06-14	X				
1010-07-01	X				
1010-07-02	X				
1010-07-03		X	X	X	X
1010-07-04	X				
1010-07-05		X	X		X
1010-07-06		X	X		X
1010-07-07		X			X
1010-07-08		X			X
1010-07-09	X		X		
1011-00-01		X			
1011-02-01		X			
1011-02-03	X		X		X
1011-02-06		X	X		
1011-02-07	X		X		X
1011-02-08		X			
1011-03-01		X	X	X	
1011-03-02		X	X		
1011-03-03	X				X
1011-03-04		X			
1011-03-05		X			
1011-03-06		X	X		
1011-03-07		X			
1011-03-08		X			
1011-03-09		X			
1011-03-10		X			
1011-03-12		X			
1011-03-13		X			
1011-05-01		X			
1011-05-02		X			
1011-05-03		X			
1011-05-04		X			
1011-05-05		X			
1011-05-06		X	X		
1011-05-08		X			
1011-05-09		X	X		
1011-07-01		X			
1011-07-02		X			

Student Code	Course 10	Course 11	Course 11A	Tweener	Different Decision
1011-07-03		X			
1011-08-01		X	X	X	
1011-08-02		X			
1011-08-03		X			
1011-08-04		X			
1011-08-05		X			
1011-08-06		X	X		
1011-08-07		X			
1011-08-08		X			
1011-08-09		X			
1011-09-01		X	X		
1011-09-02		X			
1011-09-03		X			
1011-09-04		X			
1011-09-05		X			
1011-09-06		X	X	X	
1011-09-07		X			
1011-09-08		X	X	X	
1011-09-09		X			
1011-09-10		X			
1011-09-11		X			
1011-09-12		X			
1011-09-14		X	X		
1011-09-15		X			
TOTAL /108	44	64	34	8	18

Appendix Q

List of All Unproblematic Placement Students

Student Code	Both Original & Redo Session Participant	Only Redo Session Participant
1010-01-01		X
1010-01-02		X
1010-01-03		X
1010-01-04		X
1010-01-05		X
1010-01-06	X	
1010-02-01	X	
1010-04-01		X
1010-04-03		X
1010-04-04		X
1010-04-06		X
1010-04-09		X
1010-04-12		X
1010-05-01		X
1010-05-04		X
1010-05-05		X
1010-05-07		X
1010-06-02	X	
1010-06-04		X
1010-06-05	X	
1010-06-06		X
1010-06-10	X	
1010-06-11	X	
1010-07-01	X	
1010-07-02		X
1010-07-04	X	
1011-00-01	X	
1011-03-04		X
1011-03-05		X
1011-03-08		X
1011-03-09		X
1011-03-10		X
1011-03-12		X
1011-03-13	X	
1011-05-01		X
1011-05-02		X
1011-05-03		X
1011-05-05		X
1011-05-08		X
1011-07-03		X

Student Code	Both Original & Redo Session Participant	Only Redo Session Participant
1011-08-04		X
1011-08-05		X
1011-09-02	X	
1011-09-03	X	
1011-09-04		X
1011-09-05		X
1011-09-07		X
1011-09-09		X
1011-09-10		X
1011-09-11	X	
1011-09-12		X
1011-09-15		X
TOTAL /52	13	39

Appendix R

List of All Problematic Placement Students

Student Code	Both Original & Redo Session Participant	Only Redo Session Participant
1010-01-07		X
1010-01-08		X
1010-01-10		X
1010-01-11	X	
1010-02-02		X
1010-02-03		X
1010-04-05		X
1010-04-08	X	
1010-04-10		X
1010-04-11		X
1010-04-13		X
1010-05-02		X
1010-05-03		X
1010-05-06		X
1010-05-08		X
1010-05-09		X
1010-06-01		X
1010-06-03		X
1010-06-07		X
1010-06-08		X
1010-06-09		X
1010-06-12		X
1010-06-13		X
1010-06-14		X
1010-07-03		X
1010-07-05		X
1010-07-06	X	
1010-07-07		X
1010-07-08		X
1010-07-09	X	
1011-02-01		X
1011-02-03	X	
1011-02-06		X
1011-02-07		X
1011-02-08		X
1011-03-01		X
1011-03-02	X	
1011-03-03	X	
1011-03-06	X	
1011-03-07		X

Student Code	Both Original & Redo Session Participant	Only Redo Session Participant
1011-05-04	X	
1011-05-06	X	
1011-05-09		X
1011-07-01		X
1011-07-02		X
1011-08-01		X
1011-08-02		X
1011-08-03		X
1011-08-06		X
1011-08-07		X
1011-08-08		X
1011-08-09		X
1011-09-01	X	
1011-09-06	X	
1011-09-08	X	
1011-09-14		X
TOTAL /56	13	43

Appendix S

List of All Problematic Placement Students With Their Problematic Designation(s)

Student Code	Course 11A Original Session	Course 11A Redo Session	Tweener Original Session	Tweener Redo Session	Different Decision
1010-01-07		X			
1010-01-08					X
1010-01-10		X			
1010-01-11	X	X			X
1010-02-02		X			
1010-02-03			X		X
1010-04-05		X			
1010-04-08	X				
1010-04-10					X
1010-04-11		X		X	
1010-04-13		X		X	
1010-05-02		X			
1010-05-03		X		X	X
1010-05-06					X
1010-05-08		X			
1010-05-09		X			X
1010-06-01		X			
1010-06-03					X
1010-06-07		X			
1010-06-08		X			X
1010-06-09					X
1010-06-12		X			
1010-06-13		X			
1010-06-14			X		
1010-07-03		X		X	X
1010-07-05		X			X
1010-07-06		X			X
1010-07-07					X
1010-07-08					X
1010-07-09		X			
1011-02-01	X				
1011-02-03	X	X			X
1011-02-06	X	X			
1011-02-07	X	X			X
1011-02-08	X				
1011-03-01	X	X		X	
1011-03-02	X	X			
1011-03-03					X
1011-03-06	X	X	X		

Student Code	Course 11A Original Session	Course 11A Redo Session	Tweener Original Session	Tweener Redo Session	Different Decision
1011-03-07	X				
1011-05-04	X				
1011-05-06		X			
1011-05-09		X	X		
1011-07-01	X				
1011-07-02	X				
1011-08-01	X	X		X	
1011-08-02	X				
1011-08-03	X				
1011-08-06	X	X			
1011-08-07			X		
1011-08-08	X				
1011-08-09	X				
1011-09-01	X	X			
1011-09-06	X	X		X	
1011-09-08		X		X	
1011-09-14		X			
TOTAL /56	22	34	5	8	18

Appendix T

List of Course Students' Survey of Adequacy Ranking in Diagnostic Sessions

Student Code	Slightly Below Course 10	Proto- typical Course 10	Slightly Above Course 10	Slightly Below Course 11	Proto- typical Course 11	Slightly Above Course 11
1010-01-01	X					
1010-01-02		X				
1010-01-03	X					
1010-01-04		X				
1010-01-05	X					
1010-01-06		X				
1010-01-07			X			
1010-01-08			X			
1010-01-09		X				
1010-01-10		X				
1010-01-11			X			
1010-02-02		X				
1010-02-03			X			
1010-04-01	X					
1010-04-03		X				
1010-04-04		X				
1010-04-05		X				
1010-04-06	X					
1010-04-08		X				
1010-04-09	X					
1010-04-10		X				
1010-04-11			X			
1010-04-12		X				
1010-04-13		X				
1010-05-01		X				
1010-05-02		X				
1010-05-03			X			
1010-05-04			X			
1010-05-05			X			
1010-05-06			X			
1010-05-07		X				
1010-05-08		X				
1010-05-09		X				
1010-06-01		X				
1010-06-02		X				
1010-06-03			X			
1010-06-04		X				

Student Code	Slightly Below Course 10	Proto- typical Course 10	Slightly Above Course 10	Slightly Below Course 11	Proto- typical Course 11	Slightly Above Course 11
1010-06-05		X				
1010-06-06			X			
1010-06-07		X				
1010-06-08		X				
1010-06-09		X				
1010-06-10		X				
1010-06-11	X					
1010-06-12		X				
1010-06-13		X				
1010-06-14		X				
1010-07-01	X					
1010-07-02			X			
1010-07-03		X				
1010-07-05		X				
1010-07-06		X				
1010-07-07		X				
1010-07-08	X					
1010-07-09		X				
1011-02-01					X	
1011-02-02					X	
1011-02-03					X	
1011-02-04					X	
1011-02-05					X	
1011-02-06					X	
1011-02-07					X	
1011-02-08					X	
1011-03-01					X	
1011-03-02					X	
1011-03-03					X	
1011-03-04					X	
1011-03-05					X	
1011-03-06					X	
1011-03-07					X	
1011-03-08					X	
1011-03-09						X
1011-03-10						X
1011-03-12					X	
1011-03-13					X	
1011-05-01					X	
1011-05-02					X	
1011-05-03					X	
1011-05-04					X	

Student Code	Slightly Below Course 10	Proto- typical Course 10	Slightly Above Course 10	Slightly Below Course 11	Proto- typical Course 11	Slightly Above Course 11
1011-05-05					X	
1011-05-06					X	
1011-05-08						X
1011-05-09						X
1011-07-01					X	
1011-07-02					X	
1011-07-03					X	
1011-07-04						X
1011-08-01						X
1011-08-02					X	
1011-08-03					X	
1011-08-04					X	
1011-08-05					X	
1011-08-06					X	
1011-08-07					X	
1011-08-08					X	
1011-08-09						X
1011-08-10						X
1011-09-01					X	
1011-09-03						X
1011-09-04						X
1011-09-05					X	
1011-09-06				X		
1011-09-07					X	
1011-09-09				X		
1011-09-10					X	
1011-09-11					X	
1011-09-12					X	
1011-09-13					X	
1011-09-14						X
TOTAL /109	9	34	12	2	41	11