# WHAT COUNTS AND WHY? ASSESSMENT IN TEACHER EDUCATION

By

Rebecca Ellis

# A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Curriculum, Instruction, and Teacher Education – Doctor of Philosophy

## ABSTRACT

# WHAT COUNTS AND WHY? ASSESSMENT IN TEACHER EDUCATION

#### By

#### **R**ebecca Ellis

In this dissertation I consider the ways that pre-service teachers are assessed in the middle of their program. I conducted my research at Galaxy University, a large, Midwestern university that had just completed its CAEP accreditation. Here, I collected syllabi, core assignment task descriptions and rubrics, and de-identified pre-service data submissions, as well as interviewed course instructors. I then analyzed my data to look for trends and themes, as well as with the goal to better understand the choices made around assessment decisions. Throughout my research, I paid special attention to issues of fairness and how this led to tensions in the decision making process.

Key words: Assessment, Teacher Education, Pre-service Teachers, Tensions

### ACKNOWLEDGEMENTS

### To my committee

Dr. Corey Drake – You have been my champion at MSU since before I even arrived. You took a chance on me on your research project, and soon transitioned into being my advisor. Even though my quantitative assessment drive sometimes left you confused, you stood by me and helped me to the end.

Kelly Hodges – Thank you for understanding my desires for assessment and alignment and providing me with opportunities to grow and develop. With you, I had the chance to be part of the CAEP accreditation plan, and the opportunity to influence the rubrics for TE801 and TE803.

Dr. Amelia Gotwals – You took a chance on me, joining my committee when you barely knew me, yet providing me with feedback and support all the same.

Drs. Anne-Lise Halvorsen and Margaret Crocco – Thank you for joining my committee at the last minute. I appreciate your dedication to helping me complete this project.

## To my family

My mother, Cecelia Goodman Ellis – You supported me and listened to my rants when times got tough. You never believed that I couldn't do this and have had so much enthusiasm whenever I completed a step.

My father, Charles Ellis – You are always only a phone call away. I also know that my dedication to proper grammar comes from you.

My sisters, Deborah and Jenna Ellis – You have continued to love me even though I have spent so much time away from home, and when I come home, still end up behind a computer.

i

My grandparents, Bubby and Zaidy – You support me in everything I do. I cherish the yearly Passover breaks that I take with you.

My grandparents, Grammy and Gramp – You may no longer be here physically, but you have influenced my life for the better.

### To my friends

Katie Cook – You were my reliability champion. I appreciate you taking the time to come over with Willow to help ensure that my data was coded correctly.

Matthew Feinberg – You make sure to check in with me every week (if not more often) just to say hi. You are always there for me. Thank you.

Rachel Weiss – You are always my friend and great for a story exchange. So soon we can take our post-dissertation trip!

Elizabeth Setren - You took this path first, and showed me that it can be completed.

Penguin friends forever!

Kalev Maricq – Sometimes I think you thought I could do this more than I could. You worked with me to fight through writer's block and to create a plan to finish. Thank you.

Screeemm Team (Megen, Mei, and Scott) – You put up with me and maintained my friendship even when I had to skip events, be slow at responding to messages, and at times, just needed to rant.

Nyssa Romine - What would I do without you?

Dawnmarie Ezzo - For being my best friend in the program, I thank you.

Amy Peebles – Knowing that you are always in the 116 Bay has made this doctoral experience better. I also knew I could stop by for a chat or to sit on the couches and study. Thank you for everything!

### Additional Important Thank Yous

Dr. Elizabeth Heilman – You have been my supporter and a positive light during my time at MSU. You helped me understand my own posititionality and guided me to understand what I could best research for my dissertation. Even though at the last minute you could not be on my committee, I still consider you part of the team.

Drs. Floden, Calabrese-Barton, Alonzo, and Greenwalt – You were formative members of my early advising and guidance team at MSU. I thank you.

Michigan State University – My home from 2012 until now. You took me from where I was and turned me into a doctor!

Brandeis University – My first experience in post-K-12 education, you shaped me into the person I am today. You fostered my love of learning and provided me with the tools to pursue my dreams.

Jamesville-Dewitt Public Schools – You were so rigorous that college felt easy. I was so prepared for what came next.

Syracuse Hebrew Day School – Thank you for my undeniably strong academic foundation. United Synagogue Youth – You provided me with an additional family and friends for life. Lansing Salsa Community – I would not have survived without my dance breaks!

Maricq Family – Thank you for acting as my home away from home and treating me as one of your own.

# TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
KEY TO SYMBOLS AND ABBREVIATIONS	ix
SECTION 1: INTRODUCTION	1
Chapter 1: Introduction Motivation Background of the Problem Statement of Research Problem Definition of Terms.	
Chapter 2: Literature Review Overview of literature collection Defining knowledge History of testing in the United States Fairness and bias in testing and what counts as knowledge The Myth of Objectivity Knowledge and Assessment in Teacher Education Summary	13 14 17 18 23 25 27 37
Chapter 3: Methods Methodology Sample Justification and Access Researcher's Role and Positionality Context Data Sources Data Analysis and Rigor	38 40 43 45 47 49
SECTION 2: DATA ANALYSIS Chapter 4: Overview of Cases Case 1: Course C Case 2: Course A Case 3: Course BD	59 60 63 68 72
Chapter 5: Course C Overview Syllabus Major course assignments descriptions and their rubrics Other assignments Graded PST submissions	77 77 83 90 91

Interview with Dr. Aldebaran	
Looking across	
Summary	
~,	
Chapter 6: Course A	112
Overview	112
Syllabus	
Major course assignment descriptions and their rubrics	
Graded PST submissions	
Interview with Dr. Polaris	
Looking across	
Summary	
Chapter 7: Course BD	147
Overview	147
Syllabi	
Major course assignment descriptions and their rubrics	
Other Assignments	
Graded PST submissions	
Interviews with Dr. Altair and Dr. Deneb	
Looking Across	
Summary	
S unima y	107
SECTION 3: ANALYSIS AND IMPLICATIONS	
Chapter 8: Looking Across the Cases, an Analysis	189
What can we learn about a course from different course materials <sup>2</sup>	190
What is the purpose or objective of the course according to the individual docu	ments? 198
How do teacher education courses vary <sup>2</sup>	909
How does the order of learning affect what is taught and graded <sup>2</sup>	907
What does a grade in the course tell us?	919
What the purpose for assessing dispositions, especially professionalism and rule	following <sup>2</sup>
what the purpose for assessing dispositions, especially professionalism and rule	917
Chapter 9: Tensions, Philosophy, and Implications	
Fairness in Assessment	223
How are the core assignment and the curriculum related? How does this relation	ushin
influence fairness <sup>2</sup>	995
Does subject area matter in non disciplinary courses?	930
How are dispositions factored into grading?	
What factors, other than teacher knowledge or dispositions, get assessed?	
What factors, other than teacher knowledge of dispositions, get assessed?	940
Suntharia	
Our aludia a Thoughta	
Concluding 1 noughts	
APPENDICES	

Appendix A	
Appendix B	
· ·pp ·····	_00
REFERENCES	

# LIST OF TABLES

Table 3.1 Coding Scheme
Table 5.1 Point Allocation for the Four Detailed Lesson Plans
Table 5.2 PST Description of Student Group Project
Table 5.3 PST Description of Student Reflection
Table 5.4 PPp1 Grade Breakdown for Submissions
Table 5.5 Matching PP Rubric Elements with Syllabus Objectives
Table 5.6 PP Element Weight by Syllabus Themes
Table 5.7 PP Element Weight by My Theme
Table 5.8 Matching Syllabus Objectives with How Assessed in the Course
Table 6.1 Matching Course Components with Course Objectives
Table 6.2 Transposed Matching of Course Components and Course Objectives
Table 6.3 Weighting of Essential Course Elements
Table 7.1 Summary of Key Components in PST SP Submissions
Table 7.2 Matching SP Rubric Elements with Dr. Altair's Syllabus Objectives
Table 7.3 Matching SP Rubric Elements with Dr. Deneb's Syllabus Objectives
Table 8.1 Conversion Charts
Table 8.2 Grade Bands
Table A.1 Detailed timeline of testing from mid-1800s through 1990s. 252

# LIST OF FIGURES

Figure 2.1 Complete Literature Framework	
Figure 2.2 Simplified Literature Framework	14
Figure 2.3 Domains of Mathematical Knowledge for Teaching	
Figure 7.1 SP Rubric Excerpt	160

# **KEY TO SYMBOLS AND ABBREVIATIONS**

PST	Pre-service Teacher
AACTE	American Association of Colleges of Teacher Education
NCATE	National Council for Accreditation of Teacher Education
TEAC	Teacher Education Accreditation Council
CAEP	Council for the Accreditation of Educator Preparation
NBPTS	National Board for Professional Teaching Standards
InTASC	Interstate New Teacher Assessment and Support Consortium
GU	Galaxy University
CIV	Construct-Irrelevant Variance
РР	Planning Project, the core assignment for Course C
PPp1	Planning Project Part 1
PPp2	Planning Project Part 2
C1	PPp2 Task Description and Rubric
C2	PPp1 Task description and Rubric
<b>C</b> 3	Dr. Aldebaran's Course C syllabus
C4	Electra's PPp1 submission
C5	Maia's PPp1 submission
AP	Analysis Project
ADP	Assessment Development Project
A1	ADP Task Description
A2	AP Task description sheet

A3	Kochab's ADP submission
A6	Yildin's AP submission
A7	Anwar's AP submission
A9	Dr. Polaris's Course A Syllabus
SIB	Supplemental Item Bank
SP	Summary Portfolio
ISTE	International Society for Technology in Education
B1	Dr. Altair's Course BD Syllabus Part 1
<b>B</b> 2	Dr. Altair's Course BD Syllabus Part 2
<b>B</b> 3	Dr. Altair's Course BD Syllabus Part 3
<b>B</b> 4	Dr. Altair's Course BD Syllabus Part 4
<b>B</b> 5	Dr. Altair's Course BD Syllabus Part 5
<b>B</b> 6	Dr. Altair's Course BD Syllabus Part 6
B7	Dr. Altair's Course BD Syllabus Part 7
<b>B</b> 8	Dr. Altair's Course Introduction document
<b>B</b> 9	Dr. Altair's SP Description Sheet
<b>B</b> 10	Dr. Altair's SP Rubric
B11	Okab's SP Submission
<b>B</b> 12	Alshain's SP Submission
<b>B</b> 13	Tarazed's SP Submission
<b>B</b> 14	Cyg's SP Submission
D1	Dr. Deneb's Course BD Syllabus
D2	Dr. Deneb's SP Task Description and Rubric

<b>D</b> 3	Dr. Deneb's SP Template
D4	Dr. Deneb's Fake News Task Description & Rubric
<b>D</b> 5	Rukh's Submission for the Fake News assignment
<b>D</b> 6	Sadr's SP Submission
D7	Farawis' SP Submission
SAE	Standard American English
DIF	Differential Item Functioning
ETS	Educational Testing Service

# **SECTION 1: INTRODUCTION**

#### Chapter 1: Introduction

Preparing teachers is a complicated and involved business. Search the Michigan State University online library for "preparing teachers" and you get over 1,500,000 results. In my research, I wanted to know how teacher educators know that a pre-service teacher (PST) knows what they need to know. Depending on how one defines and operationalizes "know", however, the question changes. What a PST needs to "know" can refer to understanding content, to having pedagogical skills, dispositions, and more. To "know" that a PST has this knowledge, can be assessed and understood in many different ways.

What PSTs need to know to be successful is an important question, but not the one on which I decided to focus my research. Instead, I wanted to center my study not on merely what they need to know, but presuming we already have a goal for what the PST needs to know, how do we know that they know it? I am an "assessment person" who wonders about measurement issues. I care that if we are assessing, that we are assessing what we actually want to assess, and that if we use a proxy, this proxy holds. I want to know how we are assessing PSTs, with respect to what we want to know about what they are learning. Furthermore, I want to know that the assessment choice.

While assessment is a broad topic, in this study I investigated how PSTs are measured and assessed within a teacher education program. In choosing this particular slice of assessment, I aimed to focus on a critical area where effective assessment is desired, as demonstrated by the national focus on improving teachers as a way to bolster student learning. As Ginsberg and Kingston (2014) note, "teacher preparation programs have been undergoing change for years and have embraced the idea of accountability" and give the mission statement of American Association of Colleges of Teacher Education (AACTE) and the merger of National Council for Accreditation

of Teacher Education (NCATE) and Teacher Education Accreditation Council (TEAC) to form Council for the Accreditation of Educator Preparation (CAEP) as examples (p. 7). Even in 2010, Darling-Hammond spoke of the importance of bolstering performance assessments, as they could fill the gap when "current measures for evaluating teachers are not often linked to their capacity to teach" (p. 2). And it is not just performance assessments that came to the forefront. Written tests, like the Praxis Professional Knowledge Test or the Interstate New Teacher Assessment and Support Consortium's (InTASC) Test for Teacher Knowledge (National Research Council, 2001) have also been flourishing. Nevertheless, there does not appear to exist a clear framework for which assessments to use where, when, and why at different points in a PST's development, especially one that responds to the trade-offs involved when deciding among assessment strategies. Questions arise about equity and fairness whenever a test is implemented, and yet the answers are scarcer. As I worked on this dissertation, I sought to understand how teacher educators choose and use assessments.

It was not just about choosing the right test, however. I sought to understand what claims instructors wanted to make about their PSTs, how they articulated these claims, and how they were able to determine whether the PSTs had learned enough to continue with their education and preparation. I chose to look at syllabi, core assignments and rubrics, and interview data to better understand what the instructors cared about and how they were able to decide if the PSTs were ready to move onto the next course or stage in their program. Passing must mean you know enough, right?

## Motivation

I cannot remember a time when I did not care about teaching and learning. When I was in middle school, I proposed a plan to my parents (as I was too shy to actually tell the school) where

the school would be designed around different learning styles and students could choose to take courses based on the design of those courses. Students who learned better from group projects would take the group project classes and students who learned better from listening and writing would take those classes. Even then, however, I doubted the viability of my plan. What would prevent students from choosing the classes with their friends or the ones that sounded the most fun, instead of the ones that helped them learn the most? As a middle school student, I was already questioning both how to better design schools and how to ensure that learning was a top priority.

This dedication to questioning the best way to teach and learn did not fade as I grew older. In AmeriCorps as a teaching fellow, I questioned the school's use of a pre-test. I worked under one supervisor who wanted the results from a pre-test to be used only as a baseline measure to which the post-test could be compared to measure growth, instead of to use the information from the pre-test to help tailor the curriculum to the students' individual needs. I also encountered students whose results from their test scores led to people drawing the wrong conclusions about the students' knowledge. I worked with students who seemingly struggled to understand genetics, as demonstrated by their scores on a genetics test, but I then found that the problem was not genetics at all. My students had never learned to mix colors, and thus could not explain what color the offspring would be in an incomplete dominant cross of red and white flowers, even though they could give a perfect answer using just the genotype.

As a graduate instructor, I had fears about grading within my own sections of teacher preparation courses. And it was not just about a single test, but also about how I graded and assessed my students using many different forms of assessment. These fears included those

focused on the grade output, as well as on the differential learning that might be occurring as a result of being a different instructor, which I discuss in more detail over the next few paragraphs.

In terms of actual grade giving, I feared that the students in my course might have gotten a different grade if they had had a different instructor. When sections of a course are not aligned in terms of curriculum or assessment, it is highly likely that in a different section, with a different assessment or rubric, a student could get a different final score. Unless the student is on the cut-point of passing or failing, this difference was probably not significant in the long run, but in the moment, it felt like a big deal.

Beyond a difference in score, however, if there was a different assessment or rubric, it was also quite possible that different components of the student's ability to teach were being measured. I know that when I taught TE250, a teacher education course at MSU aiming to expand future teachers' view of the world and the effects of power and oppression in society, for example, I focused on assessing my students' ability to reason through arguments and defend claims with evidence. My reasoning was that TE250 was about helping PSTs become more aware of the world around them and the inequalities, and to help them question the status quo. I did not think I could actually measure their desire to question the world, because of two very different issues that presented threats to this measurement. One issue was that it is hard to see desire in the context of one course. I did not think I had the skills to be able to assess desire since I could only use responses to writing prompts and in-class behavior as proxies. The other issue was that it could be easy to fake desire if the students knew it was being graded. Students have learned how to doctor responses to make a grade, and thus it is not always clear if motivations are genuine. Because of these two barriers, instead focused on grading their ability to make coherent, supported arguments. Another possible way to grade, however, would have been to grade based on the substance. or

content, of the argument and the degree of alignment with the ideology of the course. I chose against this method because I did not feel that I would be able to accurately measure dispositions<sup>1</sup>. If there were a change in what was written from the beginning to the end of the semester, I would not have known if it was because the student's beliefs had changed or if the student finally figured out what they needed to say to make me happy and get the better grade. Regardless, if I was grading on reasoning skills or if I was grading on content, both the grades and the claims made about the students would be very different. If there were not a standardized way to grade this across sections, the claims made about students in the different sections would consequently be not just different, but also mean very different things. Getting a B in the course, for example, would not mean the same thing across sections.

I wondered, and spent hours agonizing about this. Was one way of grading, by reasoning skills or content knowledge, more fair than the other? Was I punishing students who "got it" but could not articulate themselves well? By focusing on the ability to create supported arguments, I was privileging a particular type of knowledge. At the same time, would grading dispositions be fair? While there are arguments that certain dispositions might be better for teachers, they take a while to change and maybe it cannot be expected that one course meeting twice a week would be enough, and so grading based on disposition might have privileged those with more prior experience with the course material (Villegas, 2007). With every choice in assessment, it seemed that someone benefitted and someone suffered. This complicated my understanding of fairness, since each choice was fairer for someone and less fair for someone else. I needed to understand what it meant to be fair. Did fair mean that everyone was held to the same standard? Did fair mean

<sup>&</sup>lt;sup>1</sup>Dispositions, according to Taylor and Wasicsko (2000), are "the personal qualities or characteristics that are possessed by individuals, including attitudes, beliefs, interests, appreciations, values, and modes of adjustment" (p. 2). Assessing dispositions often includes assessment or professionalism and rule-following, and also includes assessment of attitudes and beliefs.

that everyone was able to show that they learned something in the manner that best demonstrated their understanding? Could I balance consistency from a common standard with allowing all my students to demonstrate their knowledge? Was I even capable to imagine all the ways that something might be unfair for someone? Fairness in teacher education was getting complicated, and the consequences were even further reaching.

I needed to spend more time formally understanding the choices made about assessment, especially for PSTs. I needed to know what we claim about PSTs in a given class and how we know if the PSTs meet these goals. It was too lofty to answer this broadly, so I decided to focus in on a few cases and study what happens in some classes. I decided that I would spend my dissertation looking at how select instructors for pre-services teachers in the middle of their program make their decisions. What do they want to be able to know about their students and how do they verify it?

### Background of the Problem

Choosing an assessment in teacher education is a large responsibility. This is especially challenging in a teacher preparation program when the end goal is for the students to become excellent teachers, and yet in many of the earlier courses, the instructors are not given an opportunity to observe them actually teaching children. Instead, teacher educators have to decide which proxies or behind-the-scenes skills they think they can pinpoint, augment, and then assess. There is only so much to learn about students by reading a paper or reflection. Nevertheless, it is the instructors' job to ascertain how prepared these students are to become teachers in the future. Furthermore, most instructors are not trained psychometricians and do not have a strong background in assessment decision-making or design, yet are still required to make responsible

assessment decisions. Choosing an assessment in teacher education is not only a large responsibility, but also one full of tension and controversy.

#### Tensions and controversy in testing.

Since practically the birth of standardized testing, there have been controversies and tensions. Sometimes the controversies make it to headline news, while other times they stay local (e.g. when students in Syracuse, NY confused newswoman Jackie Robinson with the famous baseball player<sup>2</sup>). Sometimes the debates are between test developers and political columnists (e.g. Lippmann versus Terman in the 1920s) and other times the debates are between other stakeholders, such as parents versus the schools (e.g. Daniel Lowen in 1980) (Haney, 1981)). Almost always, these debates fall back to the same underlying issue: how fair is the measure?

Assessment, while intended to be a fair and equitable way to determine who knows what, is still a human concept influenced by human beliefs and biases. As Haney (1981) puts it, "it seems safe to say that [assessments] ultimately are determined by social and political values" (p. 1981). Like Kincheloe (2008) argues about language, despite what some people would like to believe, assessments are not "neutral" (p. 55). Assessments measure whatever the dominant idea of knowledge is at the time and relegate people into boxes based on their results from these tests. As such, assessment is a social justice issue. Nevertheless, testing persists because as much as it is controversial, as a country, the belief continues to exist that tests provide important data and help stakeholders to make critical decisions. As Linn (2001) puts it, "the combination of enthusiastic support and strong disapproval has a long history" (p. 29).

 $<sup>^2</sup>$  In the late 1990s, there was a question on the statewide history examination that asked students about Jackie Robinson. The question developers intended to make this question about the first black baseball player, as he was part of the curriculum. However, Jackie Robinson is also the name of a black, female newscaster in the Upstate New York region. As a result, many of the students instead wrote about her. After some consideration, the state decided to award credit to either answer, since deducting points seemed to unfairly punish some students. However, either decision was controversial and led to some heated debates about fairness and education.

#### Tensions and controversy in teacher education.

This history of controversy related to assessment has a more recent aspect as accreditation for teacher preparation has come to the forefront of the national discussion around education. Norris (2013) notes that as part of the discussion around high-stakes accountability in K-12 public schooling, teacher education has also received "a heightened sense of accountability assessment" (p. 554). He argues that this is partly a response to the No Child Left Behind act that put teachers under greater scrutiny and consequently has an impact on what is expected from teacher education programs. Accordingly, more attention has been placed on what makes someone a good teacher and how it can be known that a teacher is qualified and successful. Different organizations have responded in different ways to this call. For example, the American Council on the Teaching of Foreign Languages created guidelines and assessments to better measure that PSTs were developing appropriate language proficiency before being credentialed (p. 554). As another example, in 2013, the Council for the Accreditation of Educator Preparation adopted new standards to increase the rigor and criteria for teacher education programs (Heafner, McIntyre, & Spooner, 2014, p. 516). InTASC, NCATE<sup>3</sup>, and National Board for Professional Teaching Standards (NBPTS) have also answered the call and put forth standards for how teachers, future educators, and teacher educator programs should be assessed (Kraft, 2001). Thus, it is no surprise that assessment in teacher education is a popular topic and deserves more analysis. With so many new and differing options, it can be a challenge to know which assessment to choose when and why, and the trade-offs associated with each decision may become convoluted. As stated by Cochran-Smith and Villegas (2014), "We need much more research about aspects of teacher preparation and certification-conducted with many different kinds of research designs-that deeply

<sup>&</sup>lt;sup>3</sup> Since the publication of this source, NCATE has since merged with TEAC to become CAEP.

acknowledges the impact of social, cultural, and institutional factors, particularly the impact of poverty, on teaching, learning, and teacher education" (p. 391). As this scholar points out, measuring teacher preparation is complex, and is influenced by the social structures present in our society. If we expect to be able to create fair and equitable measures, we need to also consider content and background.

### Statement of Research Problem

In my dissertation, I explore the different ways that PSTs are assessed within the middle stages of their program. I look at core assignments and rubrics, as well as interviews with professors. Before a PST enters the K-12 classroom, I want to know how teacher educators determine that the PST knows "enough."

### **Research Questions.**

- What are PSTs expected to learn, do, and know in the middle years of their teacher preparation programs?
- 2. How do teacher educators assess these middle-program PSTs?
- 3. What are the tensions involved in these expectations and assessment decisions?

## **Definition of Terms**

**Assessment** is any measurement of knowledge, beliefs, abilities, or practices that is conducted in such a way that the results are intended to be meaningful to someone. Assessments can be instructional, evaluative, and/or predictive, although Perie, Marion, and Gong (2009) caution that trying to meet too many goals for an assessment tends to lead to less effective assessments. **Course** is the term I use to describe the overarching idea behind a class on a subject. For example, Algebra I could be a course, and so could Theories of Teaching and Learning. Courses can be taught by different instructors and in different sections, but share a common goal.

A **Section** is a subset of a course and can be taught by the same or different instructor. A section has a different set of students in the course.

**Standardizing** is used in this paper to show that there is uniformity within and across something. Standardizing a test means giving the same test to all students. Standardizing sections can come from aligning curriculum, matching class activities and assignments, or having the same final. Depending on what is standardized, there are different types of alignment.

**Tension** is "a relationship between ideas or qualities with conflicting demands or implications" (Google Dictionary). In this dissertation, I use tension to refer to the competing goals in a course and how these competing goals translate to assessment strategies and types.

Equitable is a type of fairness that is measured by equal outcomes.

**Validity** means that a test score properly represents the underlying construct that it is trying to measure (Raykov & Marcoulides, 2011).

**Construct-irrelevant variance** (CIV) is when something other than the intended construct to be measured conflates with the construct, giving a student a score other than what it should be, had the irrelevant construct not been present.

**Pre-service teacher (PST)** is used in this paper to describe college students who are currently taking courses to become a teacher, or are in a student teaching program.

**Rule-following** is when a PST demonstrates that they can follow instructions. Rulefollowing is often assessed on an inclusion-exclusion scale. For example, a rubric may assess a PST

on their rule-following when it assesses **PSTs** for including 3 sources in their assignment, or using specific formatting.

Professionalism is assessed when PSTs are graded on their behavior and mannerisms.

**Dispositions**, according to Taylor and Wasicsko (2000), are "the personal qualities or characteristics that are possessed by individuals, including attitudes, beliefs, interests, appreciations, values, and modes of adjustment" (p. 2). Assessing dispositions often includes assessment of professionalism and rule-following, and also includes assessment of attitudes and beliefs.

### **Chapter 2: Literature Review**

As background to my dissertation, I researched literature in a number of different areas. I used the below framework in order to arrange and design my search.





My theory was that while my goal is to learn about measurement in teacher education (light orange box), there is considerable lead up and inputting categories that must be addressed in order to properly situate my research within the broader field. This chapter is arranged by starting with the broadest topic, defining knowledge, and then moving down the left side of the model. Once I have covered all the components that feed into assessment in teacher education, I tackle that topic directly. Under assessment in teacher education, I follow a similar pattern from the macrocosm. First, I look at definitions of knowledge in teaching, focusing first on specific types of knowledge and then looking at behaviors and dispositions. Then, I look at how PSTs are assessed, and entertain issues of fairness and bias. Finally, to look at fairness more directly, I consider both how teacher preparation programs are assessed, and how there exists variation both within and among programs. More simplistically, the pyramid below shows the direction of my research, starting from the bottom and working its way to the top, or pinnacle, of my review.



## Figure 2.2 Simplified Literature Framework

## Overview of literature collection

I first spent time researching specifically about knowledge. I chose this entry point for a number of reasons. First, much of the research on standardized testing surrounds intelligence testing, and looking further into what counts as intelligent or knowledgeable was an important first step because it would allowed me to have some background on what it meant to assess knowledge. Second, I wanted to research different conceptions of knowledge because I believed that they may foreground many of the values I would see in my cases, based on the assumption that beliefs about knowledge can influence how knowledge is assessed. Third, looking at definitions of knowledge

connected to one of the ways that I planned to analyze my collected data, which focused on how knowledge changes as the context changes.

After researching definitions of knowledge, I transitioned to researching the history of standardized assessment in the United States. Even though I am looking at standardizing on a much smaller scale, actually focusing on just assessment in teacher education, I decided that learning about how standardized testing became prominent in American society would be an important backdrop to my research. As I read, I looked specifically for tensions and for how debates around assessment persisted in different ways throughout the past century and a half (for a more complete history, see my compiled table in the Appendix). What I found was that most of the tensions presented in the timelines related to the tensions about knowledge. Questions surrounded what counts, who gets to decide, and what claims could be made.

From there, I transitioned to considering fairness and bias in testing. As testing became more prominent, so did arguments and concerns about the tests being unfair or biased. Consequently, I looked into literature about decision-making and controversies around whose knowledge counted as valuable. This shift into a more social justice stance both reflects my personal opinions about assessment and aimed to uncover the reasons for why so many people complain about tests being unfair. Furthermore, focusing on what makes tests fair or unfair was an important background to considering the tensions I found in my data.

Next, I considered the myth of objectivity. This topic built off the previous one, as it looked specifically at one reason why people find assessment to be unfair or unjust. This topic also has been prevalent in my graduate career, having been a topic of discussion in many of my teacher education courses. Because this trend was so prevalent in my teacher education experience, it seemed that focusing on this particular aspect of unfairness in educational assessment was leading

me closer to assessment in teacher education. This particular myth seemed to demonstrate why tensions were arising in assessment, especially when it came to standardizing. These tensions ended up showing themselves when I considered how the instructors I interviewed understood academic freedom.

Finally, I considered how knowledge is defined and assessed within teacher education. Since my cases all are of teacher preparation courses, it matters not just what it means to know or to assess, but what it means for PST to know and how they can be assessed. Under this heading, I looked how we compare teacher education programs, how we consider (or do not consider) teacher dispositions, and how courses can vary between instructors and formats.

Literature was found primarily using Google Scholar. I then used the MSU library to access any documents not easily freely available through Google. Some literature was found through snowballing using earlier found material. I additionally bolstered my own search terms by consulting readings that were assigned in many of my graduate classes.

To recap, in the first section below, I look at briefly at some of the different philosophical ideas around defining knowledge. In the second, I give an overview of assessment from the mid-1800s until modern day. From there, I transition into the third section where I consider literature on fairness and bias in testing. In the fourth section, I bring under consideration the idea of objectivity and how it has influenced ideas about education. Finally, I look at how knowledge is measured and assessed within teacher education.

## Defining knowledge

Assessments are often given because someone wants to know about another's knowledge, skills, and abilities. While skills and abilities can be more directly assessed (e.g. give a task<sup>4</sup> that requires doing the action), measuring knowledge is much more complicated (T. Raykov, personal communication, September, 2013). The classical definition of "justified true belief" dates back to Plato (Ryan et. al, 2002). By justified true belief, philosophers mean that to know something, one must think it is true, must think it for logical reasons, and the thing must in fact be true. This conception of knowledge, however, is widely debated and each of the major philosophical scholars has inputted their own definitions. Locke, for instance, says humans need to trust their senses since they have no other real options. Descartes (1984), however, says that humans can only have limited knowledge of absolute truth, but can reach the foundations of their knowledge by using logical inferences. Like Locke, he does contend that humans at least have to trust their reasoning, if not their senses. Empiricist Berkeley (1972), on the other hand, argues that the entire physical world is a façade, and then we cannot really know anything.

With all these competing ideas about knowledge, what does this mean for assessment in a teacher preparation program? Without a concrete definition of knowledge, assessment of knowledge is confusing and complex. My understanding, then, is that there is no best way to assess knowledge, since knowledge is so nebulous, and that measurements of knowledge are always dependent on what counts as knowledge at the given time, what type of knowledge we want to

<sup>&</sup>lt;sup>4</sup> This is more easily assessed when the skill or ability is something like "solve single-digit multiplication problems" or "be able to hop on one foot." Skills or abilities that are more nuanced, such as reacting properly in a high-pressure environment, are less easy to assess since the number of variables increases and it may be difficult to ascertain which variables impacted the examinees ability to perform as expected. In any case, measuring knowledge is still more challenging since knowledge is internal and choosing the appropriate proxy for the knowledge can introduce new complications.

measure, and the beliefs of those in charge of making the decisions. In the next section, I will build on this idea as I analyze consider the history of testing in the United States. Once I have covered a bit more background on assessment, I will reconnect back to this idea of defining knowledge when I conclude in the final section about knowledge and assessment in teacher education.

#### History of testing in the United States

While my dissertation looks at assessment in education, I begin with a history of testing in general. Some of the earliest recordings of large-scale assessment date to the mid-1800s, when Edouard Séguin used foam boards to measure the intelligence of what he considered to be cognitively impaired children (Boake, 2002). The foam board test was a test that asked children to complete a puzzle (Tulsky, Saklofske, & Ricker, 2003, p. 9). Around this time, the Digit Span Test was also used to measure memory, but the term "mental test" was not coined until 1890 by Cattell (Boake, p. 384). Also in 1890, Rice administered spelling surveys to students, and according to Haney (1981), this is often marked as the beginning of standardized testing in the United States.

In 1905, Binet and Simon published their intelligence test, and by 1908, they had modified it to include age levels (Boake, 2002, p. 386). In the 1910s, psychologists on Ellis Island developed and used mental testing on immigrants to the United States. These psychologists made their own assessments because they did not feel it was appropriate to use the Binet-Simon test booklet since that had been developed with French schoolchildren in mind (p. 388). (This concern aligns with a tension I found in my data about what sort of knowledge is being assessed in teacher education and for what purpose. Just like the Ellis Island psychologists wanted to use different items to measure the knowledge they cared about, so I assumed that current teacher educators make similar choices.) Over the next few years, additional scholars added test items and tasks to assess other areas, such as mathematics and hearing impairments. By 1916, Terman had expanded

the Binet-Simon assessment to measure adult intelligence, as well, which changed the results from being an age score to an intelligence quotient (Haney, 1981, p. 1022).

Large-scale standardized testing became first used in the United States with the Alpha and Beta Examinations for the military before World War I (Boake, 2002). These intelligence tests were administered to over 1.7 million soldiers (p. 390). Pinter (1923) found that these military tests suggested that the mental age of American adults was between ages 13 and 14. Yet, Wechsler, a psychologist working on scoring the examinations, found that the tests were often leading to soldiers who had low English proficiency receiving low mental scores despite being fully functional members of their communities (Boake, p. 394). It turned out that the test was inaccurately conflating English knowledge with mental ability (and again the tension of what knowledge is being assessed was present). Wechsler was so inspired that he went on in his career to later develop the Wechsler-Bellevue intelligence test. Thus, from the onset of standardizing tests in America, questions of fairness and the proper implementation of examinations to understand and sort people did not go unchallenged.

Tyler, a curriculum scholar, remarked that the time period from 1897 through 1927 was full of criticisms for the various assessments (Haney, 1981), although the critiques were not limited to this time period. One of the more famous debates about the assessments was between Walter Lippmann and Lewis Terman (p. 1023). The two debated publicly through their publications, with Lippmann arguing that even attempting to measure intelligence through testing was misguided, and with Terman arguing that it was foolhardy for a non-psychologist to attempt to explain things he clearly did not understand.

Despite some of the negative response to these assessments, it was not just the army administering them. News articles began suggesting that the tests be used in trade schools for

sorting purposes. In 1921, *Education Review* published an article suggesting that the intelligence tests should be used for college admissions (Haney, 1981, p. 1022). That same year, School and Society published an article on using the intelligence tests in high schools (ibid). At the same time that articles were advocating for the expansion of testing, others called for caution. Breed and Breslich (1922) published an article in *The School Review* concerned about the accuracy with which these intelligence tests could correctly place students. As evidence, they presented the difference in scores that were calculated when different intelligence tests were used. They further argued that 18 percent of students were misplaced when these assessments were implemented. Mursell (1939) was less judicious in his arguments against the testing, arguing that it was "on the level of palm-reading, bump feeling, and the casting of horoscopes" (p. 526). He further went on to argue that there is a noticeable absence of proof that the intelligence tests might actually measure what they purport to measure. Worcester and Corey (1936) also argued that the assessments are poorly supported. They gave a review of the Detroit Tests of Learning Aptitude and remarked with extreme shock (evidenced by the use of the phrase "Very likely so!") that the test was calibrated using only 50 students and was standardized against its own measures (p. 260). Nevertheless, at the same time, testing companies were working to increase their efforts to ensure that their assessments met stricter validity and reliability measures, and Tyler believed that these changes were in response to the criticisms (Haney, 1981, p. 1024).

Fast forward to the 1950s, and testing was focused on tracking and selection (Linn, 2000). With the space race and Sputnik launched in 1957, the United States was focused on being a global player and wanted to develop in science and mathematics. The National Defense Education Act, in fact, provided financial assistance to schools to administer testing (Haney, 1981). It was this

push to become a global player and to know that students were meeting the needs of this global market that led to much of the support for testing in this era (Clarke et al., 2000).

The 1960s, however, presented more backlash against testing. According to Linn (2000), this was a period of using testing for program accountability. His theory is that with compulsory education laws, policy makers wanted more oversight into what was happening in schools. ESEA, the Elementary and Secondary Education Act, enacted in 1965, included in it "demands for evaluation and accountability (p. 5). As a result, testing companies flourished. Exams were given multiple times per year in order to demonstrate gains in student achievement (p. 5). Linn also argues that testing cycles were manipulated to show the most gains. During this period, however, testing was also used for personality testing, and this drew much public attention (Haney, 1981). People became concerned about assessment result security and feared what conclusions might be made about themselves if their test scores were visible to others, especially employers. Here, the claims that could be made from testing came under scrutiny; again suggesting that what counts as useful knowledge might unfairly discriminate against some. In 1962, Banesh published The Tyranny of Testing, although he was not the only one who was opposed (Haney, 1981). A quick search of Google Scholar filtered for the 1960s with the words "tyranny" and "testing" leads to several pages of results. The New York Times posted an article in 1960 entitled "What the tests do not test" (Haney, 1981). While the contents of this article are hard to track down, the title parallels a similar article posted by the National Council of Teachers of Mathematics from 1925, written by Walker. In all of these there exists tension around what may happen to the curriculum (or even society) as tests become more pervasive, with concerns that the test could become the only real focus in education.

The distrust of standardized assessment continued into the 1970s (Haney, 1981) despite the use of tests now focused on minimum competency testing (Linn, 2000). As the tests were now used to make sure everyone was learning enough, others were worried that the tests were racist and classist. This period was marked by some heated debates between test developers and opponents about whether the tests were successful at measuring true differences between races and classes or if the tests were wreaking havoc. Truth-in-testing legislation came into effect and there were numerous court cases against testing systems and implementations out of concerns that students were being unfairly punished by biased tests. SAT scores dropped, and the College Board and ETS worked together to analyze the problem. They came up with multiple possibilities, none (of course) being the possibility that maybe the test needed revision (Haney, 1981). Instead, they blamed society, motivation, and family structure for why scores were dropping. In 1979, Lerner, the director of the National Academy of Sciences Committee on Ability Testing, lamented that there was a "War on Testing" and that this war was a cover-up for the poor teaching done by teachers and teacher preparation programs. This sentiment was echoed in 1980 by the president of College Board (Haney, 1981).

More recently, tensions surrounding assessment have been more focused on what the assessments can do and are expected to do. Haertel and Calfee (1983) worried that tests were not aligning to the school curricula, causing a disconnect. Archbald (1988) argues that tests are supposed to meet three purposes, to measure how schools and students are improving, to suggest areas of improvement, and to select which students are expected to succeed in the future. Perie, Marion, and Gong (2007) refer to these three purposes as evaluative, instructive, and predictive. Unlike Archbald who states these purposes matter-of-factly, these scholars caution, "when an assessment system purports to fulfill too many purposes—especially disparate purposes—it rarely

fulfills any purpose well" (p. 11). Perie, Marion, and Gong believe that part of the challenge of creating successful assessment is focusing on what goal or purpose is to be met and not trying to meet them all. Otherwise, the test is bound to be unsuccessful. This tension, the purpose of assessment, was one that I looked for in my data, but did not appear too often.

Despite the rocky history of testing in America, testing has managed to persist, grow, and change. Common Core State Standards, No Child Left Behind, and the Every Student Succeeds Act are all more modern implementations of curriculum and testing requirements. Because of this tumultuous history and continued existence is why I felt it was important to look more closely at the tensions present in today's teacher education environment and work to make sense of the chaos. In the next section, I transition to looking more specifically issues of fairness and bias in testing.

#### Fairness and bias in testing and what counts as knowledge

A major tension in assessment surrounds social justice and fairness. While tests are supposed to be objective and tell the administrator about the knowledge of the test-taker, time and time again researchers and community members argue that a test has made claims about a student that are not entirely true. Some of the best examples come from looking at gender. Whether the assessment is of a person's knowledge of mathematics, their violin playing, or their teaching, women and girls have systematically scored lower than men and boys throughout modern history (Ball, Cribbie, & Steele, 2013; Beidleman & Cole, 1991; Benbow & Wolins, 1996; Faber, 2008; Gallanger, Levin, & Calahan, 2002; Goldin & Rouse, 1997; Greenberg, 2010; Halpern, 1997; Loewen, Rosser, & Katzman, 1988; Navarro, 1989; O'Connor, 1992; Rosser, 1989; Sharp, 1989).

Goldin and Rouse (1997) found that when symphonies conducted blind auditions, women were more likely to be hired than when the musicians were visible to those making the selection
process during the audition. As a result, they realized that part of what made people think that a musical performance was good was attributed to the gender of the player. While my paper is not about musical testing, this case highlights the idea that what counts as knowledge, skill, or ability is sometimes attributed to features beyond what is supposedly being tested. In more technical terms, test scores are often skewed by construct-irrelevant variance, where the score is measuring something other than the intended construct (Haladyna & Downing, 2004).

According to Loewen, Rosser, and Katzman (1988) women used to score higher than men on the verbal section of the SAT and men scored higher than women on the mathematics section. Unhappy with these results and assuming that something must be wrong, the SAT verbal section was altered to include more readings on math and science, and the men's scores increased and they scored higher on both sections than women. This demonstrates that if all that was needed to show that men had higher verbal skills was to change the context, then perhaps context, and not content, was in fact influencing everybody's score. The verbal section may have been biased against men since the measurement of their verbal skills quickly changed as the passages were changed. This "quick fix" calls into question the validity of the assessment. If the scores could be so easily changed by changing the context, then how do we know that the verbal section measures verbal ability? Perhaps the verbal section is more about what one knows contextually, and this knowledge, and not one's ability to parse passages, is what the assessment is measuring. Why did no one go back and change the math section to have more word problems about material better known to women to see if that was why they were scoring lower than the men?

The controversies mentioned above highlight many of the tensions connected to testing. First, they highlight that construct-irrelevant variance can have a huge impact on a person's score. The contextual base of a problem has the ability to boost and lower scores independent of the

knowledge intended to be assessed. Second, they underline that how a test is written (by whom, for whom, with whose understanding of what it means to "know" the subject) has a huge impact on who scores well and on who we think is knowledgeable in a subject. Because the verbal section was now implicitly testing science (which was pushed more heavily on male students), men were able to score higher and the test score could be used to claim than men were better at verbal intelligence. Thus, by designing the SAT in this way, it inherently boosted male scores and supported the claim that the test-writers believed, that men had better verbal skills than women. The test was adjusted until it acted as confirmation bias.

O'Connor (1992) discusses how the success of tests, and the implementation thereof, is dependent on the goals of the assessment. Because there are many desired outcomes (such as program evaluation or student placement), what one decides to use as the examination or chooses to do with the results can vary widely. Thus, for tests to be "fair," one needs to keep the purpose of the test in mind and make this purpose known and believed by all parties. O'Connor quotes McLaughlin (1987) saying, "policy at best can enable outcomes, but in the final analysis cannot mandate what matters" (p. 9). The overarching idea is that assessments can be designed and created to give a score, but the future of education is more than a test score. Consequently, fairness will reside not in the score, but in how the people involved make sense of it. This then leads us to the myth of objectivity.

## The Myth of Objectivity

A topic that I have encountered throughout my graduate career is the concept of neutrality and how it is often looked at as being objective and correct. This idea that schooling can be removed from personal views is often a way to avoid thinking about issues of power, oppression, and bias. Applebaum (2009) discusses that the assumption that teachers can be neutral ignores the

fact that teachers and schools exist in a society and that unless the teacher actively acts to make changes, and instead maintains the status quo, the teacher is complicit in continuing a society of imbalance and unjustness. Thus, to be neutral is not to be objective and fair, but to perpetuate the systems of power in American society.

Colbert (2009) addresses this idea of neutrality in his comedic clip called "The Neutral Man's Burden." While not an academic source, he highlights the irony of white males being allowed to use their background to make decisions and still being neutral, while women or people of color doing the same thing raises the concern of bias. Similar to Applebaum (2009), his argument reminds viewers that what counts as objective is a subjective decision. Furthermore, there is no way to be truly objective, and to claim so brings in cultural ignorance or elitism.

Bringing these ideas to assessment, assessments also will not be neutral or objective, no matter how hard the assessment writers try. Questions, especially those set in a context, will include some degree of bias. A good example is the problem used by Wiest (2008) in her article on how to adapt lessons for English Language Learners. The problem involves counting 18 animals and 52 legs and then determining the numbers of pigs and chickens counted. This question, however, requires that students know that chickens have two legs and pigs have four in order to answer. Thus, this question is biased against students without this prior knowledge.

O'Connor (1992) addresses a slightly different understanding of context when she says that "judgments of test fairness are wide ranging and dependent upon context in ways that are difficult to generalize" (p. 12). By this, she explains that how "good" we claim a test to be is often dependent on how good the test results are. People are more willing to claim that a test properly measured their knowledge when they score well, and more likely claim that there was something wrong with the test when they score poorly. As such, the interpretation one takes from a grade is

influenced by what the grade actually is. This then leads to challenges when trying to come to conclusions based off a test score.

These two examples relate to construct-irrelevant variance (CIV), which I studied in my comprehensive examination. It is not just context that leads to bias, for "CIV can present itself in many ways, from measuring reading on a science item, to judging handwriting, to reflecting student motivation at the time of the test" (Ellis, 2017). CIV can also be present when those taking the test have access to different resources (Anyon, 1980). If a test score is dependent on resource access, then it is not so much measuring student knowledge but what students have had opportunities to learn. This nuance is crucial because it makes the difference between an aptitude and achievement test.

Finally, these biases often contribute to the systems of power and oppression in modern society. The system of power perpetuates because of cultural capital. Bourdieu (1973) states "Cultural wealth...only really belongs...to those endowed with the means of appropriating it for themselves" (p. 57). This means that those who get ahead in society are those who already have an advantage. As such, testing is designed in such a way that the test developers can choose who succeeds by how the test is designed (and what they determine counts as knowledge) and this keeps certain groups succeeding and in power. When the tests do not maintain the status quo, like the SAT verbal section mentioned earlier, people object and the test is changed. Perpetuating the status quo is seen as normal and objective, and challenges are called out as destabilizing and wrong.

# Knowledge and Assessment in Teacher Education

Defining and measuring knowledge becomes extra complicated when considered through the lens of PSTs, as in my cases. In this final section of my literature review, I look at assessment within teacher education from four different lenses. First, I look at understanding types of

knowledge that are associated with teaching, as in order to know what will be assessed, one first must consider what one wants to know. Under that heading, I also look at teacher dispositions and behaviors, as these are also often assessed in teacher education. Second, I look at the many different ways PSTs are assessed, both historically and currently. Third, I expand my understanding to look at how teacher education programs are assessed, not just PSTs. Finally, I look at the variation within the programs as a way to forefront some of the tensions I found in my data.

### Specific types of teacher knowledge.

Teachers need to know not just the content knowledge that they will be teaching, but must also develop pedagogical content knowledge (PCK), which is the knowledge of how to break down a topic and explain it to others (Shulman, 1986, p. 11). As part of PCK, teachers also need to have knowledge of learning and learning trajectories. Additionally, teachers need curricular knowledge, which means that the teachers know alternative ways of teaching material and what resources are available (p. 12).

Ball and colleagues (2008) build upon Shulman's framework and add some development to unpacking what teachers need to know. While their work started in analyzing mathematics, it is also fairly generalizable to all content areas. The first domain is Common Content Knowledge (CCK), which is the general knowledge of the content being taught (e.g. what is twenty times thirty, or in what year was the Emancipation Proclamation signed) (p. 399). The scholars clarify that this knowledge is not common to all people, but that it is what you would expect a teacher or other professionals in the subject matter to be able to know. The second domain is Specialized Content Knowledge (SCK), which is a knowledge special to teaching, including key components like the ability to diagnose the source of a student's error or to pay attention to nuances in the content that



#### **Domains of Mathematical Knowledge for Teaching**

**Figure 2.3** Domains of Mathematical Knowledge for Teaching (Ball et al., 2008, p. 403)

will lead to more effective teaching, such as knowing the correct next problem to give a student to help them on trajectory of learning (p. 400). The third domain is Knowledge of Content and Students (KCS), which involves being able to anticipate where students might struggle or become engaged, and to be aware of common misconceptions (p. 401). The fourth domain is Knowledge of Content and Teaching (KCT), which is what teachers know about instructional strategies and resources that will maximize their ability to help their students learn (p. 401). Together these domains (plus Horizon Content Knowledge and Knowledge of Content and Curriculum) make up what these scholars call Mathematical Knowledge for Teaching (MKT). To see how they overlap, see Figure 2.3 copied directly from the source.

# Behaviors, not just knowledge.

There is a tension in assessing teachers, and that involves measuring or assessing behaviors (often included as a component in dispositions), in addition to knowledge. For example, Lee (2005) argues that teachers need to know how to be successful reflective thinkers. The idea is that teachers need to be able to reason through why certain strategies are working or not in order to make successful decisions about their students and future teaching. Thus, in addition to MKT, successful educators know how to effectively reflect.

There is a portion of the teacher education field that focuses on teacher dispositions and how or if to measure them. According to Taylor and Wasicsko (2000), dispositions are "the personal qualities or characteristics that are possessed by individuals, including attitudes, beliefs, interests, appreciations, values, and modes of adjustment" (p. 2). As such, dispositions include who the teacher is, what they think, and how they behave. Thus, in addition to the areas of knowledge that a teacher must learn in order to become a teacher that I mentioned earlier in this review, many scholars argue that dispositions are also critical to good teaching. Furthermore, "scholars have emphasized that [teacher preparation programs] must do what they can to develop dispositions in teacher candidates" (Al-Rawashdeh, Ivory, & Writer, 2017, p. 751). Borko, Liston, and Whitcomb (2007) detail some of the key debates on including dispositions in teacher education programs. One of the key arguments for including dispositions is that dispositions are what transforms knowledge of teaching into action. This argument hinges on the belief that for a teacher to teach in a particular way, they must have an inclination to do so. One of the key arguments against including dispositions is that "dispositions cannot be measured reliably and validly" and as such there has not been empirical evidence sufficiently collected to claim that requiring particular dispositions is correct (p. 362). This debate is complex and involves many additional scholars. While I focused on PST assessments, dispositions were presented in my data, but were not the focus. Therefore, as measuring dispositions is tangential to the research I collected, this is all I will mention for now.

### Assessing PSTs.

If we want teachers to know in all these different ways (and I did not even cover that some believe that we need teachers to believe certain ways (e.g. Villegas, 2007)), then it is a safe assumption that there is not one way to assess PSTs, and that different teacher educators have different approaches. I first summarize some of the ways PSTs have been assessed over time, and then look at how PSTs are assessed in the modern era.

### Assessing PSTs throughout history.

According to Forzani (2011), normal schools in the early 1800s, which were designed to prepare teachers, did not have formal assessments for how they knew that PSTs were ready to teach. This was likely a result of the fact that early normal schools did not have codified understanding of what PSTs needed to be able to know or do. Instead, the schools were "concerned primarily with the moral and disciplinary aspects of school-keeping" and thus did not place an emphasis on how they would assess teacher skills as we think of them today (p. 25). In the early 1900s, Henry Holmes, dean of the Harvard Graduate School of Education, advocated for a written examination as the capstone of the teacher preparation program (p. 60). However, this assessment never was very successful as the faculty could not come to a consensus about the curriculum, and thus the examination remained quite broad and ineffective.

After World War II and with the start of the Cold War, emphasis shifted to strengthening education and the Masters of Arts in Teaching quickly developed (Forzani, 2011, p. 72). Programs expanded their clinical work and aimed to help PSTs develop by filming and reviewing their student teaching practice. Furthermore, attention was placed on how and what a PST should know in order to properly measure readiness. As articulated by Scates (1950),

The defining of good teaching is not impossible; it is difficult because of the psychological subtleties and because of the interplay of many factors. It cannot be accomplished in terms of any single pattern of characteristics unless these are made very general; for purposes of teacher education the objectives may have to remain on a rather general level inasmuch as we prepare teachers for many diverse situations. To assess teacher performance, however, the standards need to be more specific so as to be more observable; and they must be varied to suit the variation in local needs. (p. 141)

Thus, there was a clear focus on looking to understand what teachers needed to be able know and do and to look for ways to be able to assess this. Conclusions for how actually assess, however, were less common.

In the 1960s and 1970s, microteaching became a key way to assess PSTs (Forzani, 2011, p. 78). In microteaching, PSTs practiced and demonstrated particular teaching skills, such as questioning students or calling on students without their hands being raised. One might consider microteaching an early form of performance assessment. Additionally, some teacher preparation programs had a focus on developing portfolios that would list the competencies that a PST had achieved prior to graduation (p. 80).

#### Assessing PSTs in the modern era.

In recent years, there has continued to be a myriad of ways that PSTs are assessed. In 2009, Suzuka and colleagues wrote an article suggesting tasks for teacher educators to help PSTs develop MKT (although they did not include how to assess the success of their tasks). When I worked with Barbara Weren and Heather Howell at the Educational Testing Service (ETS), the goal was to write an assessment that would measure SCK and KCS for secondary mathematics teachers as part of the National Observation Teaching Exam. When I worked with Courtney Bell

at ETS, the project was to assess how well elementary PSTs could lead classroom discussions around topics in mathematics and English language arts using a simulated classroom<sup>5</sup>. In the field, teachers are assessed with various rubrics such as Marzano and Danielson. With all these assessments, I sometimes feel like Simon (of the Binet-Simon intelligence test) says it best with his quote, "It matters very little what the tests are so long as they are numerous" (as cited in Boake, 2002). While the quote is meant jokingly, the point remains that multiple measures are often the best way to fully understand the underlying knowledge. Assessing is hard, and there is likely no best way to do it succinctly. This becomes especially true when deciding how to measure PSTs.

Analyzing how PSTs are measured and assessed is particularly timely given the current climate around accreditation of teacher preparation programs. According to Heafner and colleagues (2014) the new standards set forth by the Council for the Accreditation of Educator Preparation (CAEP) on August 29, 2013, were designed to increase the rigor in how PSTs were admitted, trained, and graduated. Heafner suggested that a result of the CAEP changes would be that some programs would even close as a result of not meeting the new and more challenging standards.

According to the Council for the Accreditation of Educator Preparation (2016), these new standards are as follows:

- 1) Content and Pedagogical Knowledge
- 2) Clinical Partnerships and Practice
- 3) Candidate Quality, Recruitment, and Selection
- 4) Program Impact
- 5) Provider Quality Assurance and Continuous Improvement.

<sup>&</sup>lt;sup>5</sup> The test is its final form can be found here: https://www.ets.org/note/test-taker/about/elementary-education/lgd

All of these standards help to increase the rigor in teacher education programs, and are currently shaping the field of teacher education, and as such, provide a meaningful arena for my project. It is even more pressing that we know how teacher education programs are assessing their **PSTs** as this affects not only who becomes a teacher and how well school children are taught, but also because there is an implication for a teacher preparation program's future depending on the success of their assessments.

#### Assessing teacher education programs.

Additionally, there is literature on assessing teacher education programs and considering the myriad ways that variation exists both between programs and within. Darling-Hammond, Chung, and Frelow (2002) conducted a study that looked at how early service teachers (with 3 or fewer years of experience) rated their preparedness to teach. Interestingly, while there was much variation between preparation programs, responses from teachers who had taken the same path were relatively stable. This suggests that the experiences provided within a teacher preparation program likely make a difference in how a **PST** is prepared and how confident the teacher will feel in the first few years of teaching. This assumes, however, that there is not a selection bias by the program, and that it is not just that each program selects from a certain niche of people. This finding contrasts with the findings by Goldhaber and colleagues (2013), which found that there was more variation within programs than across. This, however, might be due to the fact that many programs offer several paths within their program.

Because there is so much variation between programs, there is not just CAEP looking to understand and come to conclusions. Wineburg (2006) suggested that we need to develop a "national framework for evidence" that would help teacher preparation programs collect the necessary data to present effectiveness. I would be interested to know if Wineberg thinks that

CAEP is the solution, or if something else is still needed. Also from 2006, Darling-Hammond introduces the TTK, or the Test of Teaching Knowledge developed by InTASC as a method for reviewing and assessing teacher education programs. Brabeck and colleagues (2016) advocate for using multiple sources of data for evaluating teacher education programs. As they say, we need "multiple sources of reliable and valid data" if we want to come to any conclusions (p. 165). In the style of Brabeck and colleagues, I also used multiple sources of data and attempted to ensure their validity and reliability. Furthermore, Huang and Oga-Baldwin (2015) have pointed out that much of the research on how to evaluate teacher preparation programs has come from Western countries. These scholars suggest that perhaps adding more countries could help us better learn about teacher preparation more globally. Interestingly, many of the studies of teacher education programs have looked at the use of surveys of teacher preparation program graduates to come to conclusions about how well prepared the graduates feel as a result of the program. Huange and Oga-Baldwin follow this method of using surveys. In my research, however, I did not use surveys. This suggests that my dissertation is adding to the body of research in a different way. While CAEP recommends using both survey data and other sources in Standard 4, most of the research I found used only surveys. Thus, if most of the research into teacher education has been based on surveys and self-reporting, then my analysis, which focuses on looking at the assessments themselves, should bring a fresh look at an already important field of research. What might we learn when we change the data source?

# Variation within teacher education programs.

Research has not looked only at variations between and among programs, but also between and among sections of the same course at the same university. Much of the research I found looked at the differences between online and in-person versions of the same course. Interestingly,

the results determining if the presentation method influenced learning were mixed. Kock, Verville, and Garza (2007) found that although there were differences in student scores and perceptions of the course in the middle of the semester, by the end of the course, it made no difference if the student had taken the online or in-person version. They collected their data from 70 undergraduate students at a university taking an information technology course. Carrol and Burke (2010) found similar results, that the overall perceptions of the course and student scores showed little to no variability depending on presentation format, and their study looked at MBA students studying organizational theory. Thompson and colleagues (2012) also found that online versus inperson led to similar results, and they looked at undergraduate PSTs learning about special education. "Data showed similar outcomes in both sections, suggesting that both instructional formats provide a credible means to teach content in knowledge-based courses with sections that have a large student enrollment" (p. 240). Thus, from just these three sources, which used very different data sets and collected research from different years, one might conclude that the modality of course presentation has no real effect on learning.

However, I also found several studies that found that it did make a difference if the course was taught online or face-to-face. Keramidas (2012) looked at 30 PSTs studying special education. This research looked at two sections of the same course, taught by the same instructor using a similar style for teaching. Keramidas found that the students in the online course struggled more than the in-person students. Conversely, Lancaster, Wong, and Roberts (2012), who looked at 52 nursing students enrolled in the same practitioner course, found that including an online portion to the class helped students. "Overall, students enrolled within the blended lecture delivery section performed at a level that was statistically higher than their counterparts enrolled within the traditional lecture delivery section" (p. e17). Here, modality did matter.

If modality sometimes matters, and sometimes does not, I find that it is likely not just modality that is making the difference. I would assume that how the course was taught using the modality would likely be influential. An, Shin, and Lim (2009) found that different instructor strategies with online course discussion forums changed how students act and learn. This suggests that instructor strategy then plays a part, as well. It appears that modality might be able to influence how a student learns, but that also other factors might exacerbate or mediate these changes. All in all, course variability looks like it is a topic that still has space for more research.

### Summary

In conclusion, while only a brief summary of assessment and its challenges, I aimed to highlight some of the key ideas behind assessment. Assessment depends on what counts as knowledge, and that choice is subjective. There is no true objectivity or neutrality in testing, and as such, we need to be judicious in how and when we appropriate assessments. Because of the implicit bias and power considerations, deciding how to use an assessment should require a deeper look at what is trying to be achieved and why. With every decision, knowledge will be assessed from a perspective, and as such, some will benefit and some will lose. This does not mean that assessment ought to be abandoned, but rather that we need to be more conscientious of how our choices lead to consequences, and to ensure that claims made from a test score represent only what can truly be claimed. When it comes to teacher education, as teachers are so important to the future of this country, special care must be taken with how we determine who is qualified to teach. As such, I looked at the myriad of ways we measure and PSTs.

#### Chapter 3: Methods

In this chapter, I outline how I conducted my dissertation. I explain my methodology, my sample justification, my role as a researcher, the context of my study, my data sources, and my data analysis.

I conducted my research as a collective case study, where each case was a different course in the middle portion of a teacher preparation program. I analyzed each case individually and then analyzed across my cases to look for common themes and potential areas of tension.

## Methodology

I always thought of myself as a quantitative researcher. I have a penchant for numbers and analytical thinking that has always drawn me to looking at data quantitatively. I love building spreadsheets, running calculations, and attempting to make conclusions based on the numbers. However, something was missing when I was purely quantitative. As much as I wanted to know the p-value and the likelihood of an occurrence, I also wanted to know about the nuance and the individual numbers that made up the whole. I cared about assessment, but I also wanted to know the "why" and "in what way" as I looked at and analyzed my data. I soon realized that what I needed was qualitative research.

Au (2007) claims that "qualitative research ...focus[es] on human interaction and attention to the day-to-day functioning of schools and classrooms" (p. 259). To best answer my research question, I needed to look beyond the numbers and get to know the people involved in making decisions about assessment. I wanted to know why each professor graded the way they did and know how the grade translated into meaningful claims. It was not that quantitative research was wrong or lacking, but that it was not the best means for answering my particular questions.

Researching my dissertation qualitatively afforded me the opportunity to pay attention to the nuances surrounding my topic.

### Collective case study.

Because I wanted to look closely at the nuances in the tensions within and around assessment, I chose to conduct a collective case study of courses within one teacher education program. According to Patton (2015), a collective case study required me to select key cases that would hopefully provide insight into my problem or question.

Purposeful sampling is a strategic method to identify and select information-rich cases, maximizing the most effective use of limited resources (Patton, 2015). For my project, I used a typical criteria case: a four-year teaching program, a medium-sized university, and a selective teacher education program as identified by average SAT and/or ACT scores. Within this case I focused on three courses, which allowed for the courses to serve as cases within the wider program.

According to Patton (2015), a collective case study involves selecting key cases that should provide insight into my problem or question. By looking across courses, instead of delving deeply into one and instead of finding specifically conflicting or typical cases, my collection had the potential to "yield the most information and have the greatest impact on the development of knowledge" (p. 236). As Stake (1995) puts it, "opportunity to learn is of most importance" (p. 6). In this way, I chose cases not based on "balance or variety" but based on what they would likely contribute to what I wanted to learn (ibid). Thus, this dissertation was a deliberate survey of the field with a focus on the question of tensions.

A benefit to using case studies is that they can be used to disprove generalizations, using Popper's method of "falsification" (Flyvbjerg, 2006, p. 11). In falsification, a single case is used to show that something does not work in all cases. While I was not looking to disprove something per

se, I was looking to understand the role and presence of variation and how tensions may show up when assessment decisions are made. Using the case study method also allowed me to highlight how the context mattered in each of my cases, which helped me to demonstrate that standardizing assessments may be valid only under particular circumstances. As I found that each instructor made instruction decisions that were particular to their courses, this emphasized that standardizing across a teacher preparation program may lead to incompatibilities with individual courses. Flyvbjerg emphasizes that case studies are advantageous because they "can 'close in' on real life situations and test views directly in relation to phenomena as they unfold in practice" (p. 19). As I was aiming to look at nuance, case study methodology was ideal.

I chose three critical courses in my data set, each covering different content in a teacher preparation program. All took place in the middle of the teacher preparation program, which allowed for some stability among the cases, while also allowing the difference of course content and goals to be foregrounded. I hoped that choosing courses with a mix of similarities and differences would aid me to see how PST knowledge is measured and assessed. I also hoped that by choosing these cases, I would be able to see particular tensions that might arise when making assessment decisions. Once I analyzed each course individually, I also looked across courses to determine themes and trends. The details of how this was done are described later in this chapter.

# Sample Justification and Access

In order to understand how assessment is used and understood within teacher education, I chose to look closely at three courses at my target university. I met and interviewed four instructors about three core courses that all PSTs at that university needed to take in the middle of their program. In addition to the interviews, I also collected syllabi, core assignment descriptions and

rubrics, and de-identified student work. The specifics of why I chose each course are detailed in the "Sample Context" section.

While at first focusing on the middle years was the result of how my recruitment worked out (these professors were the most willing to meet with me), I soon decided that focusing on the middle years would allow me to better focus my research, minimize variables, and find generalizable data. What it means to be a PST - and to assess a PST - changes as an individual gets closer to becoming a certified teacher. In the beginning of my site's teacher preparation program, PSTs take content-focused classes from those departments, as well as education courses on human development and child psychology. My understanding is that at this point, idea of teaching is still abstract and the courses focus more on developing theory and understanding. In the middle of the program, there is a shift to developing particular skills and techniques, as the PSTs learn the how's and what's of teaching, such as how to develop a lesson plan or student assessment. Near the end of the program, the PSTs start interacting more with students and get a hands-on understanding of the profession though student teaching. Once in this third phase, assessment of progress often comes from multiple different sources (mentors, field instructors, university professors) and the picture of what it means to succeed is quite variable depending on the source. Therefore, by focusing my research on the middle years I was able to:

- 1. Focus tightly on how university professors determine adequate progress,
- Be less worried about varied background experiences that may influence progress (in the first year of a program, background experience may more play a part in what the PST knows, but by the middle, the PSTs have developed a foundation of shared experiences and knowledge),
- 3. Look at written assessment, rather than contend with performance assessment, and

4. Consider courses that all PSTs take or knowledge that we expect all PSTs to have before entering the classroom regardless of subject area or grade level.

This fourth point is perhaps the most compelling reason for my dataset. As I looked at courses in curriculum development, assessment design, and use of technology, I noticed that all three were courses that cover critical skills for teachers. Whether someone plans to teach high school biology or third grade social studies, they are expected to do what is taught in these middle courses. Thus, I found that using this dataset, that of mid-program courses, afforded me the opportunity to look at a subset of teacher education that all PSTs experience in some fashion. How PSTs were assessed in these courses would be indicative of how all PSTs are assessed in the middle of their program. Therefore, I felt that my initial recruitment responses lead to perhaps the best dataset I could analyze to answer the questions I had about teacher education, and would lead to the most generalizable findings.

I originally anticipated that decisions around assessment would be mainly tied to course aims, but I quickly discovered that it was much more involved. By adding in analyses of the syllabi, core task descriptions and rubrics, and sample work to what I could learn from the interviews, I was able to get a broader understanding of what was really being assessed and how. The interviews provided a good overview, but only ended up being one component.

I also wanted to separate out my experiences from those of these cases. Because I have experience teaching in a teacher preparation program, I wanted to be extremely clear about what it was I was learning versus what ideas I already had with me. Thus, before I collected my data, I wrote about my personal experiences in teaching (using an auto-ethnographic style) to separate out what I knew and believed coming in. I then had these journals to use as reference in case I needed them.

Getting IRB approval for this project was fairly straightforward. The professors I interviewed were all adults, and thus only needed a basic consent form to participate. The work I had from students was all de-identified or publicly posted on the Internet, so I did not need extra consent for that.

# **Researcher's Role and Positionality**

I conducted my research as an interviewer with some insider knowledge. Although I have never taught at my target university, I have taught in a teacher preparation program, so I have some idea about the structure and expectations for PSTs in the middle of their program. This insider knowledge allowed me to focus more explicitly on how assessment occurred in my sub-cases and to spend less time learning the ins and outs of a teacher preparation program. This is not to say that I did not spend any time learning about the particulars at this university, but merely that developing the sense of how the program worked took less time and effort.

As I conducted my research, I had to remember that just because I care about assessment and alignment, that is not necessarily the focus of all university course instructors. Part of why I wrote my auto-ethnographies was so that I could leave my own feelings about assessment aside as I interviewed and analyzed, knowing that I could come back to my own opinions later. With this separation, I was able to be a better active listener and learner as I gathered the data.

I have a number of personal beliefs about assessment. I start with a quantitative position, and from there I believe that results need to be not just quantifiable, but generalizable and informative. For me, if students get a B in the course, I should be able to translate that B into some sort of meaningful conclusion about their learning during the course. I also want to be able to break down grades and assessments into components, because I believe that a portfolio is oftentimes more meaningful than a straight score. I believe strongly in using grades from

assessments as diagnostics (or "instructive" as Perie, Marion, and Gong (2009) would call it). Rather than being focused on evaluating a student or predicting what they can do, I want to use my assessments to determine what the student knows and what the student still needs to work on. In fact, when I teach, almost every assignment can be revised and resubmitted because I care more about the learning than knowing that the student could do it right the first time. As I interviewed my participants and reviewed my collected materials, I was mindful that I take this approach with my own assessments, that this approach may not be commonly shared, and attempted to clarify with my participants how they expect to use assessments. I needed to look at how their assessments align with their own goals for the assessments, not just with my own.

I also believe that assessments need to be both culturally relevant to the students and consistent with what the course is trying to teach. My comprehensive examination was centered on validity and construct-irrelevant variance, and thus I am particularly sensitive to ensuring that scores from assessment truly reflect the underlying knowledge of the student. When I assess, I try to look for potential biases in my assessments and worry about the potential for a student getting a question wrong because the context of the problem was unfamiliar. When I taught a multicultural teacher preparation course, I left space in my curriculum and in my assignments for students to bring in and examine their own beliefs. For me, this met the goal of meeting the students at their level, being culturally relevant to their experiences, and also keeping with the purpose of the course to get students to stretch their understanding of the world and inquire about the injustices that surround them and their future students. The challenge, however, was how to grade these assignments. Should I have been grading the PSTs on understanding the content, or should I have included the PSTs beliefs that what I taught was true and that they had an inclination to work toward change. This question was especially difficult because there is a debate currently brewing in

MSU's teacher education program about whether that course should be measuring dispositions, which mirrors the broader question about measuring dispositions across teacher education as a whole.

As mentioned in the introductory chapter, I personally gave credit if students analyzed and used data to support and question their assumptions, but I know that other instructors felt compelled to grade based on what was said and concluded, rather than how the argument was framed and supported. Here is where a tension about knowledge came into the limelight, as even within the course instructor group there were disagreements about what counted as knowing the course material. A tension of alignment and standardization was also present because had we standardized our rubrics across the sections, these rubrics would have come into conflict with the differing ideas of what the course was meant to do. As a teacher educator, I encountered specific tensions that led me to consider the purpose of training teachers.

I understand that not every instructor will come to the same conclusions that I did when I was teaching, and as such, I used my interviews strategically to understand how and why my research participants came to their conclusions. As I interviewed the course instructors, I looked for areas of tension and questioned my participants about how they made assessment decisions based on how they viewed these tensions.

# Context

I collected my data from a medium-sized university in the Midwest of the United States. I reached out to a number of Teacher Education Programs, and it was from this university (pseudonym: Galaxy University) that I received the most replies. A benefit of working with this program was that it had recently gone through CAEP accreditation, so the professors in the program were already thinking about assessment, alignment, and quality education (not that

instructors at other universities do not think about these concepts, but here, it was at the forefront of everyone's minds).

The education department at Galaxy University (GU) is located in the old library of the university, and maintains the feel of studiousness. Next to the education building is the building for the College of Health and Human Services, which feels appropriate, as both education and health are necessary for human development and success. Located on a hill, the windows in the Education Building look out to a marvelous and expansive view.

I used three courses in the program as my cases. Each took place, as mentioned previously, in the middle years of the teacher preparation program. All took place after the foundational education courses, but before disciplinary methods courses and student teaching. I provide a brief each course below. The specifics of each course will be described in more detail in Chapters 4-7 of this dissertation.

# Case 1: Course C.

Course C is a curriculum course for PSTs in the middle of their program. In this course, PSTs learn how to develop both individual lesson plans and full unit plans. This course is interesting since it focuses on the mechanics of how to plan, but since it comes before disciplinary methods courses, the PSTs do not yet have the specialized content knowledge or knowledge of content and students needed to build their lessons. As such, the focus of the course is more structural, by which I mean that the courses focus on what is a course objective or what should be the format of a jigsaw lesson. Additionally, there is some attention to general practicality, by which I mean that PSTs are taught to consider how these structural components should fit into a larger curriculum. Dr. Aldebaran, course instructor, discussed how it was beneficial to have this course prior to the disciplinary methods courses because the disciplinary instructors could expect that the

PSTs had the foundations developed and then could focus more on content. This course was interesting to study because it focused on lesson planning without the PSTs actually having to teach the unit, and thus what made the unit "quality" was dependent on the instructor, and not on students who might one day learn from this unit.

## Case 2: Course A.

Course A is an assessment course for PSTs in the middle of their program. In this course, PSTs learn about the basics of traditional tests, collect exam data and analyze it, and write their own summative assessment. This course also focuses on the structural component of exam writing, as PSTs do not yet have disciplinary methods knowledge. This course was interesting to study because PSTs were assessed on their ability to design summative assessments and measure student knowledge before they had the skills and experience to know how to teach the content at a developmentally appropriate level.

# Case 3: Course BD.

Course BD is a technology course for PSTs in the middle of their program. In this course, PSTs learn how to use technology to enhance student learning, as well as learn when and how it is appropriate to use technology. This course afforded me extra opportunities to learn about assessment choices as it was the only course in my set that had more than one course instructor teaching the sections, and had a section that was entire online. With this additional course structure, I was able to learn even more about the decisions around assessment.

### **Data Sources**

To answer my research questions, I used several types of data. When I scheduled my interviews, I also asked my participants to give me copies of their course syllabi, core assignment task descriptions and rubrics, and de-identified **PST** submissions to these assignments. The

reasoning for my choices of data sources can be found in the next chapter. Some of this data was provided to me in advance, and I was able to review it before interviewing the instructors. Other material was given to me during the interviews, and the professors described them to me as they handed them over. Other instructors did not have the materials in advance or at the time, and sent them to me much later.

I also interviewed instructors in all three courses, using a self-designed interview protocol. Before my interviews, my advisor reviewed my protocol and approved it. Then, on the day of my interviews, I used the protocol as a guide for the conversation, taking notes on the protocol and audio recording the conversation for future transcription. A copy of the protocol can be found in the appendix. The interviews were semi-structured, and as such, the instructors and I followed the protocol fairly closely, but also deviated when the conversation dictated. As I interviewed, I realized that I needed an additional question that explicitly asked about fairness in assessment, and added it to the protocol.

After I conducted my interviews and collected the data, one of my committee members suggested that I get one more set of additional data sources to make my understanding of the courses more robust. An initial idea was to get some classroom observations to help develop a more ethnographic understanding of my contexts. On further review, however, my advisor and I decided that the best additional data source would be a graded assignment that was not the core assignment. The reasoning for this was that it would provide a more concrete understanding of what is included in the course grade, in addition to what I learned from the interviews and syllabi. Furthermore, as this assignment would be personalized to the instructor, it would allow me to see how the course was individualized by the instructors. I then reached out to my participants and asked them to provide me with a task description of an additional graded assignment, the

corresponding rubric (if they had one), and a de-identified PST submission of this assignment. Based on how their courses were designed, I received different types of responses. The exact details of these sources are explained in the following chapters.

# Data Analysis and Rigor

I completed three types of analysis for this dissertation: document analysis, interview analysis, and cross-case analysis.

### Document analysis.

As I reviewed all the submitted documents, I spent time analyzing and consolidating what I was learning. All the documents came directly from the course instructors that I interviewed, and as such, I reviewed the documents both with and without the lens of the interview.

While reading the documents several times, I used grounded coding to look for what I could learn about expectations, grading, and claims. I detail the questions I was looking to answer below.

### **Expectations:**

- What directions are explicitly given to PSTs for how to complete assignments?
- What comments or corrections are written on the submitted documents that indicate what was good or missing in a submission? What can be inferred from these comments about what is expected of PSTs?

## Grading:

- What is said in the rubrics and in the syllabi about how PSTs will be assessed?
- What is being assessed? Content? Structure? Something else?
- What does it mean to get an A on an assignment? A C?
- What does one need to be able to do to pass this course?

Claims:

- What is the purpose of the course according to the syllabus? The assignments?
- What can we know about a PST once they complete the course? What can we expect them to be able to do?

While the above questions were what I was looking to understand, I spent time also understanding in general what was given and stated in the documents. As I read each document, I used topical coding and took notes on what stood out to me. I took notes both on the documents themselves, highlighting key components, marking thoughts and interests, as well as creating separate annotated documents for some of the submitted work, where I went paragraph by paragraph and noted both what was included in the document and what I could infer from it. I worked iteratively through the different courses, and as I found something new, I would look back at the previous courses to see if there was evidence for it there, too. For the syllabi, I both highlighted components, as well as did word count checks to determine what was important and at what worth.

As I built my chapter overviews, I combined the grounded and topical coding together, building a template to describe and explain what I had found.

### Determining reliability.

As I coded my individual cases, I wanted to be sure that my coding was reliable. In this section, I explain how I determined reliability for measuring word counts and for analyzing the Course C syllabus.

# Reliability of the word counts.

As mentioned above, I highlighted the syllabi to determine relative word count. I wanted to know how strongly issues of professionalism, rule following, and use of general mechanics came up in the syllabus with respect to the total syllabus. To ensure that I was highlighting correctly, I asked a trusted colleague, Katie Cook, to help me run a reliability test. To do this, I gave her one of my four syllabi and asked her to highlight it, looking for the same concepts as I did. Once completed, I counted how many words she highlighted and calculated it as a percentage of the total syllabus. Katie highlighted 41% of the total text. When I had completed the same task, I highlighted 42% of the text. We did not highlight the exact same things, but it was quite similar. Thus, I was able to determine with confidence that I was measuring the quantity of professionalism correctly. *Reliability of analyzing the Course C syllabus.* 

I also asked for Katie's help with determining that I had grouped Course C's rubric elements correctly. In the Course C syllabus, there are fifteen course objectives divided into five themes. As I read the objectives, however, I felt that there was another way that they could be grouped that might provide some insight into what matters in the course. I particularly wanted to find objectives that aligned to designing and creating curriculum, since that is in the course description, but was not listed as a separate theme. Thus, I tried to create a group for curriculum, and then used open coding to sort the rest.

To check that I had grouped these bullets correctly, I created a matching chart for Katie with the fifteen objectives down the side and my category names across the top. I briefly explained to Katie what my category names meant, and then I asked her to fill in the chart. She put a 1 whenever there was a match and left the cell empty for no match. Once she was done, we reviewed the document together to look for comparisons and discuss any disagreements.

For this chart, there were 71 cells to compare. When considering all 71 cells, we agreed 93% of the time. When only considering agreement in the cells where I had originally marked yes in my document, we had 83% agreement. Where we differed was with what we included as being

included in designing curriculum. Katie was stricter in what she felt should be included, while I had been a bit more lenient, as I figured a course on curriculum should be able to connect many objective bullets to it on principal. What this meant for my research, then, was that the links I found about how the course connected to curriculum were even more tenuous than I had originally claimed. Using Katie's matching scheme, the course objectives are actually 27% about curriculum designing not 40%. As a trade-off, Katie found the course to be 33% about considering students, not 20% like I found, and 13% about engaging in controversial issues, not 7%.

### Interview analysis.

In addition to reviewing the documents, I wanted to see what the instructors said about the course in their own words. Using my notes on my interview protocol and from the transcripts, I reviewed each interview to see what stood out as important to each course instructor. I wanted to know answers to the following questions and coded using grounded coding:

- How did the instructor describe the purpose of the course?
- How did the instructor describe the different grade bands in the course?
- How did the instructor make decisions of what to teach and assess in this course?
- How did this instructor's course match or differ from other sections (in their own words)?

In March, right after my interviews, I wrote a memo on what I had learned in the interviews and to keep track of my initial thoughts and understanding, which was based off of the topical codes I discovered through early readings of my notes and the transcripts. I arranged this document around a few categories, namely, who I interviewed, general trends, and notable differences between the interviews.

### Cross-data analysis.

In addition to looking at and analyzing my data individually, I also looked across my data to conduct a deeper meta-analysis. I conducted analyses both within each case and across cases.

## Within each case.

Using the data I collected and my analysis, I attempted to build matrices of what was being assessed, how, and with what weight. Individually with each document, I was able to see how components were assessed, but I wanted to develop a full picture of assessment and claims. With each matrix, I looked at what the rubrics stated was being assessed, and then matched it to the claims of the course. Together, I was able to get a better sense of how the components of the course came together to make a grade. Additionally, I mathematically calculated in the weights of each rubric element relative to the total course, and included that in my chart. I then used the results of this matching, combined with the relative weight, and was able to calculate approximately how much weight each course objective earned during the course of the semester (or within a core assignment).

I also looked substantively within each sub-case to determine major themes and trends. I wrote course overviews combining the data from my collected documents and from the interviews. I started by considering only two categories, what the course was about and how the course was assessed. I soon realized, however, that I needed to break these categories down even smaller.

About the course:

- What are the course's pre-requisites?
- What is the content of the course?
- What is the general structure of how the course is designed?

Grading and Assignments

- What is the core assignment?
- What else, besides the core assignment, is included in the course grade?
- What constitutes an A in the course?
- What constitutes a sufficient grade for passing the course?

Thus, I then used the information I had gathered from all my sources to consolidate them into these categories and to answer these overarching questions. Looking to understand this helped to guide my cross-case analyses.

### Determining reliability.

In order to determine reliability for my matrices, I asked my colleague, Katie Cook, to review my process. In preparation, I took random samples from my matching tables and recreated them in a separate file for Katie to review. The theory was that a random sample should be representative of the full tables, and should be sufficient. I created one table for matching a random sample of rubric elements Course A's core assignment with the course objectives listed in the syllabus. I created another table matching a random sample of Course C's core assignment rubric with the fifteen course outcomes listed in the syllabus.

Before Katie completed each matching chart, I gave her a brief overview of the course and the core assignment. I also provided Katie with the task descriptions to use as reference. A few times, Katie asked me clarifying questions, as using a random sample removed some of the rubric elements from necessary context. After Katie completed all the matching charts, we worked together to compare results and come to a consensus.

For the Course A document, there were 91 cells to compare. We agreed on 95% of all the cells with respect to matching. As an additional check, I looked to see how many of the cells that I marked as "yes" Katie did, too. Using this calculation, we agreed 86% of the time. The major place

we differed was with how we considered what it meant to meet the objective of being able to explain the basic principles of assessment. Katie included, and I did not, PST written work about designing tests. I had thought of these descriptions as either being rule-following answering questions or about constructing a test. Other than that, we had near perfect agreement.

In the Course C document, our reliability was not as strong. When considering all the cells, we agreed 92% of the time, but when considering only my yesses, it was down to 68% of the time. As we discussed, it looked like that for one row I had actually made an error in which cell I had selected (it was completely unrelated so I had likely read my rows wrong), and as a result, I went back and updated my document. Another area we differed was in how many cells we would select when a rubric element was about considering students. Because there were so many syllabus bullets for that, I had often chose the one I felt most related, but Katie was more relaxed and chose multiple. The other major difference was with our understanding of curricular content and learning outcomes. Katie matched course concepts with content and measuring skills with outcomes. I had linked concepts with outcomes. As a result of this discussion, I reviewed my coding for this document and updated my findings accordingly. The analysis chapters that come later will reflect the changes made from this discussion.

Overall, I found that my reliability was sufficient enough to ensure that the claims I made in this dissertation had merit. With most of our matching being in the 90%-range (and only lower when I looked for specifics), I felt confident with my coding choices.

### Across cases.

To look across my cases, I based my template on Au (2007). Like Au, I used "thematic metasynthesis" that developed as I read across my analyses of the individual cases (p. 259). As part of the analysis, I created codes in multiple iterations. The first iteration was developed using the

literature and my initial read through (using grounded coding). I listed, with the help of my advisor, what questions I expected to be able to find and we worked together to develop these questions into broader ideas. These questions were:

- 1. What can we learn about a course when looking at only one data source (e.g. just the syllabus or just the interview)?
- 2. What can we learn about a course when we look at data sources together as a set?
- 3. How might issues of personal relationships influence assessment scores?
- 4. What is the role of content and content-specific ideas when assessing non-content practices?
- 5. How does professionalism factor into the assessments? Why might professionalism or rule following be graded?
- 6. Do assessments focus on measuring the type of educator the PST is becoming or do they focus on how PSTs educate?

As I coded my full data set, I looked for how I might be able to answer these questions. Additionally, I used open and emergent coding to find themes and then added these new themes to the schema. I read each of my paragraphs looking to understand what major category was covered. I made a spreadsheet with the following headers:

Table 3.1 Coding Scheme				
Major Theme	Minor Theme	Instructor/Course	Page	Evidence

I started with filling in the instructor or course code and then with the page number of the chapter where I was looking. Then, under the evidence column, I wrote a brief description of what was being described in the paragraph. If the topic lasted for more than one paragraph, I combined them together for one code, but if the paragraph covered more than one piece of content, I split it into two rows. Then, I went to the minor theme column and gave my evidence a name that summarized what I was finding, such as "instructor choice" or "variability." Once I had a full course chapter with these columns filled in, I read back through and gave each minor theme a major theme. I then repeated the process with the next course chapter. Once I had established all my themes, I then went back through all my data and made adjustments, as necessary, as well. Once my data was fully coded, I sorted my rows alphabetically first by major theme and then by minor.

Using the spreadsheet of my open codes (themes), as well as the list of my grounded questions, I built a few matrices in Excel of what I knew and could find from my data. One matrix had major themes down one side and the course instructors across the top. A second had collected materials down one side and the course instructors across the top, with a title of "what can we learn from...?" My third matrix had a similar format as the second matrix, but the cells were answered with the focus question "what is the purpose of the class according to...?" As I combined the lists or coding and questions and as I used them to fill in my matrices, I also went back into my source material to find what else was needed to fill any empty cells.

Before I started my analysis, I had expected to be looking at three specific tensions, namely questions about what counted as knowledge, how assessment influences curriculum choices, and areas for alignment. However, when I actually encountered my data, I found that starting with open coding was more reflective of the data I collected and thus adjusted accordingly. The themes that stood out became:

- What can we learn about a course from different course materials?
- What is the purpose or objective of the course, according to the individual documents?
- How do teacher education courses vary by instructor and over time?

- How does the order of learning affect what is taught and graded?
- What does a grade in the course tell us?
- What is the purpose for assessing dispositions, especially professionalism and rule following?

Once I had established that these were the questions I would be able to answer, and my matrices were fully developed, I wrote my looking across chapter using the matrices as a blueprint. Then, I realized that I also wanted to focus on tensions that arose, not just questions, so added a final chapter to focus upon that. The tensions for this final chapter came from discussion I had with my advisor, as well as themes that stood out to me from reviewing my data. This chapter is more philosophical, and as such, uses the data I collected and analysis I conducted as a springboard to talk about lasting impressions and ideas for future research. The topics I use for the tensions chapter area:

- Fairness in Assessment
- When and how should teacher education programs assess English language skills?
- How does a core assignment influence the curriculum and fairness?
- What happens when we require access to resources?
- Does subject area matter in non-disciplinary courses?
- How are dispositions factored into grading?
- What is in a rubric and how does that shape the assessment?

**SECTION 2: DATA ANALYSIS**
# Chapter 4: Overview of Cases

Determining what counts is a tricky process, and begins with delving into the definition of "counts." One common definition is "matters." Thus, to ask what counts in a course is to ask what matters and the purpose of the course. However, "counts" also has a more numerical definition, and in this sense, it can be asking what will count toward the grade. In this section and dissertation, I will be looking at both definitions.

There are several ways to determine what matters in a course, and where one looks for answers may lead to different results. Therefore, in this section, I will be looking at four major indicators of what matters in a teacher education course:

- 1. The syllabus,
- 2. Major course assignment descriptions and their rubrics,
- 3. Graded PST work, and
- 4. Interviews with one or more of the current professors.

Each course has a syllabus that is given to the PSTs to help them understand the purpose of the course and how they will be assessed. In order to understand the first official impression PSTs have of a course, and to understand the outline for a course, it is best to start the analysis of this course with this document. My goal in analyzing syllabi was to review these documents to understand what is expected of the PSTs both during and by the end of the course.

At Galaxy University<sup>6</sup>, all courses in the Teacher Education Program include at least one major course assignment, often referred to as the "core" assignment. This assignment remains the same across all sections of the course, and is designed by the department. I was not given an exact understanding of how this was developed, nor who was involved in the writing process, but only

<sup>&</sup>lt;sup>6</sup> Each course will be based upon a constellation, and have a number code attached to it. This fits with the theme of the university code name: Galaxy University.

learned that it was an agreed upon assignment and that each education course has one. This assignment is meant to ensure some consistency across different sections and terms of a course, and to maintain reliability about what the course is attempting to teach and assess. PSTs are required to pass this assignment with at least a C, in addition to passing the course itself. How the core assignment is implemented, what exact instructions are, and how it is graded are still up to the instructor, but it nevertheless provides for some consistency. Thus, the core assignment acts to create a common assignment across all sections of a course, while still allowing for a bit of instructor influence. For each course, I look closely at the core assignment task description and its rubric.

For some courses, I also was able to gather additional information about other graded assignments. When that was the case, I also analyzed this data.

In addition to looking at assignment task descriptions and rubrics, it is helpful and important to look at PST graded work. What an assignment looks like from a written description, and how it is presented when turned in is not always the same. Some of this difference may be due to how the task was discussed in class. Whatever the case, it is often helpful to look at sample work to see how the rubrics were implemented and to better understand what the assignment looks like in practice. As I reviewed PST submitted work, I looked with a lens to see what could be known about PST knowledge as a result of the assessment.

Additionally, I had the opportunity to talk to instructors of the three courses that I studied. For each interview, I used an interview protocol that I developed ahead of time. These interviews were helpful in bringing to life the syllabi and the other collected documents. I had the opportunity to probe the professors about their thinking and reasoning behind their teaching and assessment

choices. I include the analyses of my interview notes and transcripts to add to the understanding of what is assessed in each course and why.

Finally, I conclude each case with a section looking across all four data sources. After analyzing each source individually, I built a matrix of claims and assessments to construct a model of what counts and how much.

As I coded and decoded across these data sources, I aimed to understand what was important for PSTs to learn in each course and how this learning was measured and assessed. My analyses are not intended to be evaluative, but are instead to gain a deeper understanding of what is happening and why in relation to assessment of teacher candidates. Much of the current research in teacher education focuses on student teaching and performance assessments, and so my goal was to help expand the understanding of what comes before these final portions of teacher education. Thus, I coded to understand what is happening in middle courses within a teacher education program, and aimed to highlight how teacher knowledge is measured before a PST is ready to perform.

In my descriptions of the courses and the assignments, I have renamed the tasks and paraphrased some of the assignment directions. This was done deliberately to maintain the confidentiality of my research participants and to keep their course materials secure.

As an overview of the next three chapters, I provide case summaries here. In the next several chapters, I will go into detail about my analyses of the different documents that I gathered for each course.

## Case 1: Course C

# About the course.

Course C is a course for secondary education majors that focuses on curriculum and methods. Course C (or the elementary equivalent) is a required course for all PSTs at GU. This course happens in Phase 2, which takes place after the general education courses. PSTs take this course as a pre-requisite or co-requisite to the assessment course (Course A). While pre-service elementary teachers have had some experience with curriculum design already, for secondary PSTs, most of the material is new. As described by the professor, this is a "zero to sixty course."

### What are the pre-requisites?

The professor, Dr. Aldebaran, expects that the PSTs will already have learned about adolescent development and know about children. They need to know about learning theories and motivation. These expectations come from the courses that are required before taking Course C. Additionally, PSTs have also already taken a course that focuses on special education and inclusion, so those topics are not really covered in Course C, either. By the time they take Course C, PSTs also have already declared what subject(s) they intend to teach once they graduate.

# What is in this course?

The course begins with an overview of the foundations of teaching. PSTs discuss, but do not write, their teaching philosophy. However, through journals and class participation, they are encouraged to develop what they consider to be their philosophy. At the end of the semester, they write this more formally. They also engage with their starting dispositions and decide where they need to grow most. As Dr. Aldebaran described it, "all of that is happening in the first week because you have to know yourself first." From here, the course transitions to understanding the students, looking more specifically at learning styles and learning preferences. Dr. Aldebaran only gave me a brief overview of the curriculum, and thus I did not get specifics about what happens in this section of the course.

Then, the focus shifts to a major topic of the course: planning. Dr. Aldebaran commented that the first thing the PSTs will likely see in the field is a curriculum map. Therefore, in her course, she brings in maps from all the subject areas for the PSTs to analyze. These are not abstract maps, but actual curriculum maps used by teachers in the local schools. This analysis is coupled with discussions about Twenty-First Century Skills and Bloom's Hierarchy, since from here the direction is to build learning objectives.

Dr. Aldebaran explained that by this point in the semester, she is not only teaching but also modeling. For example, **PSTs** work in jigsaws and other student-centered methods so that they do not only hear or read about these teaching methods, but are able to experience them first hand. This method of hands-on learning continues into the next section of the course, in which the **PSTs** learn about multiple intelligences. As part of learning about inquiry, the professor has the **PSTs** do a group investigation and then teach what they learned to their peers.

As the semester starts to near the end (and the work has been ramped up), Dr. Aldebaran switches the class time to be more relaxed. She brings in a guest speaker to discuss connecting with the culture and community of the schools. She also takes her PSTs on a field trip to a local school that exemplifies using Twenty-First Century Skills within the curriculum. The course also becomes more of a hybrid course, with online discussions about the material. In class, the PSTs are reviewing each other's lesson plans and curriculum (which is part of the core assignment).

Then, once their core assignment is turned in, the course winds down more with an openbook online multiple-choice exam based on the course text. This is included because Dr.

Aldebaran has found that it ensures that the PSTs read the course material. At this time, the PSTs also complete a number of reflections. One reflection is on the process of their major assignment, and another is on their dispositions.

# About the grading and assignments.

To receive a passing mark and earn credit for this course, PSTs must meet two goals. They must receive a C or better in the entire course, and they must earn a 73% or better on both parts of the curriculum major assignment, the Planning Project (PP). The PP is a departmentally designed assignment that remains a constant requirement for this course no matter how the course is taught and by whom. This dual passing mark is a way that the university ensures that all PSTs passing through this course meet the necessary requirements, and is part of their accreditation. While there is much freedom in how the course is taught overall and how assignments are weighted, by having this core assignment with a minimum passing score, the university maintains some alignment and control.

### What is the core assignment?

The PP, at least with the professor I interviewed, is completed in two phases, the Part 1 (PPp1) and Part 2 (PPp2). In our communication, Dr. Aldebaran commented that she implemented the assignment in phases because she found that the PSTs performed better when the large project was broken into more manageable pieces. She also found that providing a template not only made the assignment easier for the PSTs to complete, but also made grading more uniform and simpler. Thus, while she splits the project into two phases, not all instructors teaching this course would split it in this way, or at all.

As part of PPp1, the PSTs have to choose a unit goal, find and list the corresponding student standards, and propose an assessment plan. While this course is a pre-requisite or co-

requisite to the assessment course, Dr. Aldebaran does not expect a strong summative assessment, but a more general understanding of how and what the **PST** hopes to assess.

In PPp2, the PSTs set up a lesson sequence and plan, design at least three lessons in their entirety, and develop an annotated bibliography. In our discussion, Dr. Aldebaran noted that sometimes other instructors require that the PSTs complete several more lesson plans, but she felt that focusing on just a few achieved the same goal. She also required that one of the lessons be a jigsaw, but that was not a general course requirement.

Interestingly to me, while the general core assignment stays constant across all sections of this course, the individual course instructors design the exact instructions and the rubrics for the components. Each instructor has the freedom to adjust the core assignment (as long as they maintain the essence of the assignment) and to weight the components of the project according to their own interpretation of the assignment. However, Dr. Aldebaran did not think that these changes would be major and that, overall, one could still make the same claims about the PSTs exiting each section.

# What else is included in the course grade?

Because passing the core assignment with a C or better is necessary in order to receive credit for the entire course, the Dr. Aldebaran weights this assignment heavily in her overall course grade, with approximately 50% of the overall grade coming from this assignment. This is not, however, required weighting. For the other 50%, Dr. Aldebaran assesses classwork, group presentations, field experience (and the related journals), a disposition reflection paper, a reflection on the unit planning process, a final demonstration of learning, an exam based off the course text, and professionalism. When asked why she grades in this way, the professor explained that she was "heavy on reflection." She also mentioned that she needed to keep professionalism in the grade,

because otherwise she found that (from semesters when this was not included) PSTs exhibited behaviors that did not match her expectations. The one component that the Dr. Aldebaran does not grade is the peer review.

### What constitutes an A?

I asked Dr. Aldebaran what a student would need to do to get an A in the course. In response, she described a PST who demonstrated they had "put the time in" (transcript). This would be visible with a strong connection between the PPp1 and the PPp2. An A student would also demonstrate effort with a detailed annotated bibliography, as in here the PST demonstrates that they have "really done their research" of how their chosen topic has been taught by other instructors (transcript). To the professor, this sort of detail would exhibit that the PST was putting in the effort to think deeply about their lesson planning.

### What is sufficient?

I also asked about the line between passing and failing in this course. Dr. Aldebaran described the passing line as where the PSTs are meeting the minimum requirements. This looks like a lesson sequence that holds together, with learning objectives, but not fully filled out. While these PSTs may have missed some learning here or there, they can "get by." The professor also told me that she allows for revisions and scaffolds support. She is available for appointments throughout the semester, and also has a peer review of the curricula in class before the due date. PSTs who earn less than a C on their major assignment are afforded the chance to revise and resubmit. This resubmission can earn up to a B. PSTs who do not get below the C may not resubmit.

# Case 2: Course A

# About the course.

Course A is a required assessment course for all PSTs at Galaxy University. This course happens in Phase 2, which takes place after the general education courses (which typically take two years to complete). PSTs take this course after or concurrently with the curriculum course (Course C). While the professor I interviewed, Dr. Polaris, acknowledges that assessment is a vast and varied field, as this course is an introduction, he focuses primarily on academic tests. The course is fairly technical, as PSTs learn how to write a traditional test, how to execute alternative tests, how to use these assessments for teaching and learning, and learn about current high-stakes policies related to testing and assessment.

There are five main goals for this course:

1. Be able to explain basic principles of K-12 student evaluation and assessment

2. Be able to meaningfully critique tests and other assessments

- 3. Be able to construct quality tests and other assessments
- 4. Be able to analyze and use assessment data effectively

5. Advance their levels of professionalism (Syllabus, 2017).

# What are the pre-requisites?

The professor expects that the PSTs will already have developed their philosophy of education. Before entering his classroom, the professor expects that the PSTs have already considered and have views of what they expect their future classroom to look like. He also expects that they understand the demands of teaching and are prepared to balance the various roles that teachers often embody, like "mom, dad, grandma, grandpa, good cop, bad cop, social worker..." (transcript).

# Why teach this course?

Because Dr. Polaris was my contact at GU, I ended up having a bit of a longer conversation with him than I did with the other professors. In this time, he told me about his personal connection to Course A. Dr. Polaris has been teaching this course for many years at the university, and as the program has shrunk (as many education programs have shrunk), he is often now the sole instructor for all sections of this course. His personal belief, which comes not only from his experiences at GU, but also as a K-12 teacher and administrator beforehand, is that "better approaches to classroom assessment can go a long way to promoting better teaching and learning" (interview). As such, he aims to help his PSTs learn to use assessments to create mutual respect between students and teachers, because good assessments can act as a positive conversation between student and teacher. He is also motivated to teach this course because of the political atmosphere, which places emphasis on high-stakes testing.

# What is in this course?

The course begins with some broad ideas about assessment to help the PSTs understand what the course is about. These are concepts like what is assessment, what are definitions of assessment, et cetera, which do not take long to cover.

Next, the course turns quickly technical, as the PSTs dive into the technicalities of a "quality traditional test." In this unit, PSTs learn both about what makes a good test, as well as practice writing tests themselves. They learn both from critiquing, as well as learning a list of "do's and don't's" (interview). As Dr. Polaris puts it, writing a good test is a "skill," and so PSTs learn to do it well by practicing.

In the next section, the course shifts to data analysis. The purpose here is for PSTs to know not just how to write and give an assessment and to be able to grade it fairly, but also know how to

make sense of the results and know what needs to change moving forward. The PSTs spend time doing detailed item analysis. As part of this section, the PSTs are tasked with finding a current classroom teacher who will allow them to administer a test for them, and then grade and analyze the results (this is the Analysis Project). To help them analyze, Dr. Polaris gives them a model to follow. (This assignment will be detailed more in a later section).

Before the course fully shifts to the next major topic, there is a brief interlude to focus on learning and development. There is some attention given to Bloom's and other taxonomies, as well as paying attention to cognitive load. While not explicitly stated, I inferred that this was to help the PSTs to not only write their own assessments, but to also help with the analysis of test items and results.

From here the course shifts again, this time to alternative assessments, such as project and performance. Not too much time is focused here (as those are more embedded in the curriculum course), so the course quickly shifts back to high-stakes assessments. They consider what makes a test standardized, and how tests can be norm-referenced to a bell curve or not.

Finally, the course concludes with a unit on grading. This section focuses on the variety of ways that the PSTs may have to grade in their future classrooms. As the specifics of how to grade will be dependent on where the PST is hired, Dr. Polaris keeps this unit more open to considering multiple ways, instead of giving a particular method. Included in this unit, Dr. Polaris covers topics such as weighting, grading homework, and considering classroom behavior in the grading scheme.

When Dr. Polaris teaches this course, he opts to work from a course pack instead of a textbook, because he finds that he often disagrees with some of the rules put forth by the authors and spends too much time dealing with contradictions. Instead, he uses a course pack that he updates regularly to meet the needs of the course and the changing scholarship.

### About the grading and assignments.

To receive a passing mark and earn credit for this course, students must meet two goals. They must receive a C or better in the entire course, and they must earn a 70% or better on the Assessment Development Plan (ADP). The ADP is a departmentally designed assignment that remains a constant requirement for this course no matter how the course is taught and by whom. This dual passing mark is a way that the university ensures that all students passing through this course meet the necessary requirements, and is part of their accreditation. While there is much freedom in how the course is taught overall and how assignments are weighted, by having this core assignment with a minimum passing score, the university maintains some alignment and control.

### What is the core assignment?

To help the PSTs complete their ADP, Dr. Polaris provides the PSTs with a detailed task description that covers what needs to be included in each of the ADP sections, suggested page lengths, as well as a description of the grading schema. The ADP is designed to go along with a real (or pretend) unit that the student designs (or hypothetically designs). The PSTs must create an assessment blueprint, design the assessment and provide an answer key, accommodate the test for a student with a special need, create an alternative assessment to measure the same knowledge, and reflect upon the process.

# What else is included in the course grade?

While the ADP is a critical component of this course, it is not the only way that the PSTs are assessed. The Analysis Project (AP) (mentioned earlier when discussing how they learn about grading) is worth 15% of the grade. The AP is also a departmental requirement, although does not have the same 70% pass requirement. There is also a midterm and a final. The professor also reserves 10% of the grade as a "movement toward professionalism" (syllabus). While the professor

acknowledges that this final aspect is a bit subjective, he finds that it is a necessary part of his grading scheme.

### What constitutes an A?

When asked what was required to earn an "A" in Course A, Dr. Polaris described a PST who has received high marks on all components of this course. Dr. Polaris does not offer any extra credit, because he believes that this would destroy the alignment between the assessment and the curriculum. Nevertheless, the professor aims to have the course grade reflect mastery of the material. Before due dates, PSTs are encouraged to turn in drafts and meet with him to discuss ideas before submitting a final assignment.

### What is sufficient?

I also asked about the line between passing and failing in this course. The professor described the passing line as getting the 70% on the CAP and in the course. He explained that 70% would be that the student "learned enough." In his 18 years of teaching, only two PSTs have ever failed his course (which excludes those who dropped the course before failing). The PSTs are, however, able to retake the course and their new grade would completely replace a previous fail.

# Case 3: Course BD

### About the course.

Course BD is a required technology course for all PSTs at Galaxy University. This course happens in Phase 2, which takes place after the general education courses (which typically takes two years to complete). This course typically comes after Course A and Course C, and before disciplinary methods courses and student teaching. For this course, I interviewed two instructors, Dr. Altair and Dr. Deneb. Dr. Altair teaches an online version of this course while Dr. Deneb teaches a hybrid version. The purpose of this course is to help PSTs "critically and creatively" employ technological techniques in their future classrooms (B1, p. 1).

There are seven main goals for this course:

1. Become a lifelong learner of technology for teaching,

2. Become a leader for supporting student empowerment,

3. Be able to "inspire students to... responsibly participate in the digital world" (B2, p. 2),

4. Be able to collaborate with colleagues and students,

5. Be able to design authentic learning experiences,

6. Be able to facilitate learning through technology, and

7. Be able to use data to make decisions.

## What are the pre-requisites?

The professors expects that PSTs will already have basic technology skills, such as how to use a computer, search the internet, and create a text document. Essentially, they expect that the PSTs know how to use technology for personal use, but do not know how to use it for teaching. PSTs also need to have already taken Course C and Course A, and therefore need to know about curriculum and assessment.

# What is in this course?

This course is meant to give PSTs experience using technology to enhance student learning and to improve student understanding of content. PSTs participate in a number of modules and activities that develop their understanding both of technological tools and how and when to use them appropriately.

For the previous two courses, I detailed an overview of the order of the course. However, for Course BD, which has multiple instructors, the structure varies from section to section. As

such, instead of giving a detailed overview of the two sections I studied, I instead here describe what is common to both.

In both courses, the material covered includes creating a webpage, a WebQuest, and a learning management system. PSTs engage with ethical issues surrounding technology and do a group project on an issue of their choice. PSTs also research new and interesting technologies and present them to their peers. They design lessons using different technological tools, and learn to problem solve using spreadsheets. While the actual assignments vary from year to year and from instructor to instructor, the general purpose is to get PSTs thinking about how to infuse technology into their everyday teaching practice.

# About the grading and assignments.

To receive a passing mark and earn credit for this course, students must meet two goals. They must receive a C or better in the entire course, and they must earn a C or better on the Summary Portfolio (SP). The SP is a departmentally designed assignment that remains a constant requirement for this course no matter how the course is taught and by whom. This dual passing mark is a way that the university ensures that all students passing through this course meet the necessary requirements, and is part of their accreditation. While there is much freedom in how the course is taught overall and how assignments are weighted, by having this core assignment with a minimum passing score, the university maintains some alignment and control.

# What is the core assignment?

The SP is web-based summary of how the PST has met the seven standards of the course. A PST must create a multi-page website that not only explains in prose how they have met the goals, but also provides examples of how and where in the semester they demonstrated the standards. This is an experience of both being reflective on individual growth, but also showing the

technological skills that have been covered. **PSTs** embed live hyperlinks to their other work they have read or written to demonstrate how they understand and have met the standards.

Additionally, PSTs are expected to design their website to look professional, which is explained as being free from spelling and grammar errors, and using a design that "is significant and contributes to the overall communication of professionalism" (B10, p. 3). There is no further explanation of the design on the rubric.

To help PSTs succeed with their submission of the SP, both instructors provide a detailed template for the core assignment to help clarify the expectations. Additionally, Dr. Deneb allows PSTs to write a sample reflection and turn it in early for feedback, as this helps the PSTs know if they are on the right track for the SP.

# What else is included in the course grade?

While the SP is a critical component of this course, it is not the only way that PSTs are assessed. Instructors of this course assess the other assignments and provide feedback to help the PSTs learn. Professionalism is also assessed, as late work loses points and PSTs are expected to work together in a collegiate manner. How these other assignments are chosen and weighted depends on the instructor.

Dr. Altair has fourteen modules in her course, and each one has a graded activity that goes along with it. All assignments are "project-based," as she believes that using technology in education cannot be adequately assessed with a test. Dr. Altair describes her grading requirements as, "To receive credit for an assignment, the artifact you create must be **technically** correct **AND** have appropriate **pedagogical content**" [emphasis from original](syllabus). Thus, she requires that her **PSTs** submit work that is reflective, responds to the prompts, and uses materials and content correctly. Dr. Deneb uses a point system for grading. Each assignment is worth a particular amount of points, and then at the end of the semester, these points are converted into a grade. Dr. Deneb weights the Moodle task and the core assignment the most, while other assignments are worth less. She also grades PSTs on their contributions to online discussions.

### What constitutes an A?

What one needs to earn an A in this course depends on the instructor. For Dr. Altair, an A means doing what was expected. This means that they demonstrate understanding of course material and complete assignments well. For Dr. Deneb, an A also depends on how professional the PST has been and that the PST has shown a willingness to grow and learn more.

# What is sufficient?

I also asked about the line between passing and failing in this course. To fail the SP, PSTs need to either get below 63% or 69% depending on the instructor. To fail the course, the PST must not be able to demonstrate that they can do the material nor that they understand what they should be doing.

# Chapter 5: Course C

"It's a zero-to-sixty course...It is general. It's overwhelming...The project really becomes kind of the core focus." – Dr. Aldebaran (transcript)

## Overview

Course C is a curriculum course for PSTs in the middle of their program. It comes after the general education courses, but before subject-area methods courses and student teaching. It is also the earliest course of the three I studied for this dissertation. Course C is a pre-requisite or corequisite course for Course A, and a pre-requisite course for Course BD. Thus, anything taught in either of those two courses cannot be expected of the PSTs in Course C yet. Course C is taught by Dr. Aldebaran most of the time, and Dr. Aldebaran was the sole instructor in the semester when I collected my data.

For continuity, all documents related to this course are marked by the letter C and a number. All people associated with this course have names attached to the Pleiades constellation, with Dr. Aldebaran (the key star in finding this constellation) as the professor.

## **Syllabus**

The syllabus for Dr. Aldebaran's course begins with a "note" about field experience. All PSTs enrolled in this course are expected to not only attend class and complete assignments, but also have a field placement. This note does not mention how many hours<sup>7</sup>, merely that they exist. That this note starts the syllabus appears to highlight that it is something worth considering early on.

Next, there is the course objective for the course. It is that the PSTs be "caring professional educators for a diverse and democratic society" (C3, p. 1). It is important to note the broad nature

<sup>&</sup>lt;sup>7</sup> I learned in the interview that the requirement is 30 field hours.

of this course objective, and that it has nothing explicitly to do with developing (at least structurally) a curriculum. Instead, the purpose of this course is more about the kind of teacher that the **PST** becomes and the type of education they will be able to provide future generations of students.

Once the note and objective are stated, the rest of the first page is divided into five "themes" (C3, p. 1). Each theme is numbered and bolded, and then followed by a bulleted list of things the "Teacher Candidate will be able to do" (ibid). All in all, there are 15 bulleted items that the PST is expected to be able to do by the end of the semester. Each item starts with a verb (such as "demonstrate" or "accommodate"), showing that these claims are not about passive knowledge, but about skills and abilities (ibid). There is no explicit description about how each bullet will be assessed (although I will make connections in the Looking Across section later in this chapter). Each theme appears to be connected to being caring and/or professional.

It is notable that across the five themes, none of them use the word "curriculum" or "lesson planning". Instead, the focus is on general knowledge and dispositions needed for being a successful teacher. Theme 1 is about being committed to all students learning, Theme 2 is about being knowledgeable of "content, pedagogy, and educational technologies", Theme 3 is about being reflective, Theme 4 is about professional dispositions and communication, and Theme 5 is about diversity and democracy (C3, p. 1). This indicates that while the course may have curriculum in its title, much more is important to this course. Dr. Aldebaran is looking to develop well-rounded future teachers who care about compassion, diversity, and are student-centered in their thinking. I categorized the bullets by their major idea, the method for which determining this and being reliable with my choices is described in the methods chapter. I found that of the fifteen bullets, only six mention anything about designing lessons, and one of these six is really more about understanding student backgrounds and how to address them when planning. Four of the

bullets discuss modes of communication (such as how to talk with parents or peers), three are about considering students (plus the one mentioned above), one is about transitioning from being a student to a teacher, and one is about a "willingness" to engage in controversial issues (C3, p. 1). Thus, it appears that while curriculum is the name of the course, it is only explicitly covered in approximately one- third of the course content.

Next in the syllabus is the list of required course materials. PSTs are to acquire two texts, and have an online subscription to see other important readings. PSTs are also encouraged to bring an electronic device to class, as the syllabus mentions that class time will occasionally be allocated to doing work. However, in the policies section later in the syllabus, PSTs are warned that using their electronic devices in class in such a way that decreases their active participation in class will lead to decreased professionalism points for their grade.

Dr. Aldebaran then includes a numbered list of "instructional methods and activities" that will be used in the course over the duration of the semester (C3, p. 2). Separated from the assessments, as described next, this section is about the methods Dr. Aldebaran will use to help the PSTs learn the material. If there were more than one instructor for this course, it might be interesting to see if the different instructors were using different methods to achieve the same ends.

In the middle of page two of the syllabus, Dr. Aldebaran describes and explains the assessments and the grading process for this course. While the types of assessments are listed, point values and/or weights are not mentioned. Thus, PSTs can see that there will be a midterm, course papers, draft submissions, participation grades, class discussions, smaller written assignments, the Planning Project (PP), a presentation, a self-reflection, and a peer evaluation, but it is not mentioned how the grades for these components will work together to make an overall

course grade. They are, however, given a number to letter conversion chart, so that it is clear, for example, that a 93 in the course would translate to an A- (C3, p. 2).

At the bottom of the page, there is a brief explanation about the core assignment in the course, the Planning Project (PP). The explanation notes that the assignment will be composed of two parts, PPp1 and PPp2, worth 100 and 400 points respectively. There is no mention of how many points there are total in the full course, although Dr. Aldebaran said in the interview that she likes to weight the PP to be about half the course, so it is likely that the full course is worth about 1000 points. PSTs are told that they must receive a 73% or better on both components of the core assignment in order to pass the course. Both parts of the task may be resubmitted (by a certain date) if the initial target grade is not achieved. This indicates that the purpose of this task is not to necessarily meet the goal on the first try, but through participation in this course and careful revision, success may yet be possible. While it is not explicitly explained, it is mentioned that PSTs will need to get approval on PPp1 before being able to continue with PPp2, and that whatever is approved in PPp1 must be incorporated into PPp2. If the PST wants to alter the plan for PPp2, they must first revise and resubmit PPp1.

There is a short list of bulleted points that accompany the PP overview. These bullets pertain to the importance of following directions and doing what is expected in the PP. First, if there are changes to the PPp1 after it has been approved (and new approval has not been attained), then the PST will receive a zero on the core assignment. Second, if a critical portion of the PP is missing, then the PST will receive a zero on the core assignment. Third, if the project is not turned in in a binder with "appropriate section dividers," then the assignment will receive a zero (C3, p. 2). Thus, there are three ways to fail the core assignment (and, consequently, the course) for a failure to properly follow the directions and complete the required components. While there are

three ways to fail the assignment, however, the ways differ in how substantive they are. The first way to fail looks at the big picture of the assignment, the second way depends on missing a critical component of the assignment, and the third way is purely based on not following instructions. Therefore, all three ways depend on rule following, but emphasize different components and goals of the course.

Page three of the syllabus is mainly dedicated to course policies. The first policy attends to participation and attendance. Dr. Aldebaran tells students that coming to class every time is only worth partial credit. To receive full credit for attendance, students must also exhibit "active engagement" (C3, p. 3). Like as will be seen in Course A, participation is critical to the course grade, and the definition of what this looks like in practice is not fully explained. Second, there is a note about general professionalism. PSTs are expected to be on time to class and remain the full time, and this point is emphasized by underlining the note. In the interview, Dr. Aldebaran mentioned that professionalism had been a problem in previous terms and that she had been forced to add professionalism to the grading to get better classroom behavior. It is likely that this is how she was trying to fix this problem. PSTs should not be using electronic devices, as this will detract from their participation (and the participation grade). Also, if they are to miss a class, it is the PST's responsibly to get missed work and notes and to make up any assignments. Thus, as in all three of my cases, professionalism matters. Third, there is a note about assignment submissions. The title of this note is underlined for emphasis, and then the entire note is in italics, demonstrating its importance. This note discusses that all work, other than the PPp2, will be submitted electronically. It also says that all work must be "written to professional standards" and that "points will be deducted for mechanical errors" (ibid). This point again highlights the rulefollowing component of the course and reinforces the idea that professionalism, and not just

development of content knowledge, will matter for this course. The fourth policy is about plagiarism. Dr. Aldebaran tells her PSTs that citations must be used when taking work from other sources, and that the university-wide policy will be applied if the PSTs plagiarize. Interestingly, despite the third policy on assignment submissions, there is an additional fifth policy on writing standards. In this policy, in addition to discussing the rules for spacing (which are bolded and underlined for emphasis), Dr. Aldebaran adds that assignments submitted with "many technical errors" cannot receive any higher than a C, which, from the conversion chart, is a 73-75 (ibid). It is not explained if the assignment will be graded and then lowered to a C if necessary, or if the C is the starting maximum and then additional errors will be deducted from that. There are also not any references for where to get outside help in improving writing skills, although she does provide some strategies for how to edit one's own work (such as being aware of commonly confused words, and rereading an assignment "several hours after completing it" to be able to review it with fresh eyes and find any errors (ibid)). Finally, there is a sixth policy on due dates. "Late assignments," which is bolded, will lose 50% of their grades, which is italicized, and influence the professionalism grade (ibid). Thus, there is a grade attached to timeliness that influences not only the specific assignment grade, but also another component of the course grade. It appears that timeliness with assignments is highly valued.

In summary, as part of the syllabus, Dr. Aldebaran puts forth a number of things that will be assessed in the course. There is the list of fifteen bullets that the PSTs need to be able to do, although how they will be assessed is not yet made explicit. She provides a list of assignments, without weights or grades, preparing the PSTs for what they will need to do and on what they will be assessed during the semester. She also highlights areas where students will lose points for not following directions or being unprofessional. And, she discusses the core assignment. Thus, the

syllabus gives a starting point, but not many specifics, about how all the components will come together to make a grade by the end of the semester.

#### Major course assignments descriptions and their rubrics

The major assignment for Course C is the Planning Project (PP) where PSTs need to develop a full "problem-based or project-based" unit that could be taught in their future classroom (C2, p. 1). It is divided into two parts, PPp1 and PPp2. As stated above, PPp1 is worth 100 points, and PPp2 is worth 400 points. While they are both components of the same major core assignment, Dr. Aldebaran requires PSTs to score at or above 73% on each part, and so I will discuss them separately in this section. Before I divide them, however, it is important to note that the rubrics for these core assignments are embedded into the task descriptions. From what I was given from the professor, it appears that students are to use the rubrics (and commentary) as a guide for how to develop and format their submissions.

# PPp1.

PPp1 has PSTs come up with a plan for their unit. For this part of the core project, PSTs are to complete a given template in order to describe and explain what they hope to accomplish in their unit. However, when looking at the rubric on the next page, it becomes apparent that this template is only part of what needs to be included in the PPp1. Thus, without the directions that were likely given in class, how to use this template is a bit confusing. Nevertheless, one can expect that PSTs are encouraged to read through the entire task description document, including the rubric, before beginning.

Before completing the template (at least according to rubric order), PSTs are told to justify their choice for their unit plan. They need to explain why the project that will guide the lesson is "meaningful and relevant" for the students (C2, p. 2). How this needs to be justified appears to be

open ended, although it likely is related to what was discussed in class. Also, it is not just any students for whom the lesson must be meaningful and relevant, but "today's" students (ibid). This suggests that there is a temporal importance for the authenticity, and that PSTs cannot just choose a unit because it would have been meaningful for them when they were students. This focus connects to the syllabus where there is an emphasis on understanding the students. This rationale is worth 15 points; however, how these points are allocated is not explained.

PSTs need to list both the grade level that they will be planning for, and a title for their unit. It is not clear what the purpose is of creating a title, but it is necessary to include regardless. Having a topic that matches the grade, however, is worth 5 points, so perhaps it is not so much about how the title is written, but what it conveys about the unit and its appropriateness for the students that matters. From there, PSTs need to give the goal for their unit. To understand what needs to be included in the goal description, it is necessary to reference the rubric. The rubric says that the goal must be "written in global terms and links learning and Standard" (C2, p. 2). Including the goal described in this way is worth 15 points, and is tied for the highest point value of any individual component, so it suggests that this is critical to the course. Next, PSTs are to list the appropriate Michigan standards that connect to the unit, and then to listing the knowledge and skills that the students will learn in the unit. While these two elements are listed separately, when I looked at how submissions were evaluated, it appeared that these two were graded together. Listing Michigan standards is worth 5 points and listing the knowledge and skills is 10 points, but essentially PSTs were graded on both together out of 15 possible points.

Once the PSTs have developed their rationale and filled out the template, they are asked to develop an assessment plan. Since this course is a pre-requisite or co-requisite to the assessment course, PSTs do not have much experience with developing assessments, and thus are likely

graded with this in mind. It appears that the assessment is supposed to be a project, as describing, "what students will do in project" is the first bullet in the rubric for this component. Seeing as the opening line in the task description was about a project-based unit, it makes sense that the project would be the assessment. That this description is "clear" seems to be what will be needed to get the 10 points. There is not a component in the rubric dedicated to the quality of the project. PSTs get another 10 points by listing the "content criteria" for the project and the alignment of the project to the content goals. Thus, PSTs need to show how they will align the assessment to the purpose of the unit. For another 10 points, PSTs get credit for describing the "criteria for product," although from this document alone, it is not clear what differentiates these criteria from the content criteria (but it does become clearer when looking at PST submissions). Then, for five points, PSTs need to list any formative assessments they will be using in the unit, and for another five points, how they will have students reflect upon the unit. Most of the points awarded in this unit are for listing, and it appears that PSTs do not have to give much rationale for their choices.

Finally, PSTs are required to include a bibliography, and this provides the last 10 points in the project. To receive these points, PSTs need to use at least one textbook, one standards webpage, and five other sources. Thus, there are seven necessary resources in order to get 10 points. It is not clear how these points will be divided.

# PPp2.

PPp2, which comes after approval on the PPp1, has PSTs actually design the unit that they described in PPp1. As a description, Dr. Aldebaran writes on the assignment sheet, "Lessons and activities need to be structured to meet the Unit Goal and prepare students for success on the performance assessment" (C1, p. 1). Thus, as PSTs work on transforming their unit into a reality, they need to keep in mind both the original goal and the final assessment that they will administer.

Dr. Aldebaran gives a detailed structure for PSTs to follow. It starts with an overview of seven sections that need to be included in their final product. The first section is a summary of the PPp1. The second section is a "lesson sequence" of what will be included in the unit. Then, PSTs need to create four detailed lesson plans, one lecture, one jigsaw, one group investigation, and one guided discovery, each given their own section. Finally, PSTs need to submit an annotated bibliography.

Like with the PPp1, the task description is mostly embedded in a detailed rubric. Other than the general overview and the list of required sections, all other written instructions are provided within the rubric itself. To analyze this document, I went through the rubric to ascertain what is necessary to include and how much each component weighs (in points). The total PPp2 is worth 400 points, bringing the full core assignment, the PP, to 500 points.

As a general note about the rubric, each element in the rubric has a point value, but not a gradient or scale. There is a general comment at each mini-rubric that says, "Must minimally address the following for a 'C' grade" (C1, p. 2). It is unclear then what must be done in order to get full points, or how the points are allocated within each element. This gives **D**r. Aldebaran freedom for how to grade, but leaves it unclear to a reader of the rubric. There is also no explanation of how weights or number of points have been selected for each element, although presumably it is tied to the general importance of each piece.

At the top of the rubric, there is a yes/no check box for inclusion of a signed peer review. This suggests that PSTs are required to get a peer to read through their project before submission, and that they are likely supposed to address some of the concerns presented by their peers. This component does not have a point value attached, however, so it seems to be more about the process than the actual review. Also for zero points, but on the rubric, is an inclusion of a title page. PSTs are expected to have a title for their unit, list the grade for their unit, and write their name. While this title page, and the peer review mentioned above, have no points associated with them, there is a professionalism grade that is worth 2 points, which is 0.5% of the overall PPp2 grade. To get these points, one must use dividers and follow the format. The point value is small, but the fact that it is written near the beginning of the rubric indicates that while following these rules may not really influence the content of the unit, it is still helpful, at least to Dr. Aldebaran, to have all the submissions follow the same format. Furthermore, by following the format, PSTs will likely help themselves score well, as it will likely help them to keep track of all the components and be sure that they did not skip anything.

Next, there are 100 possible points for creating the lesson sequence. The elements for this component are bulleted, but not weighted. These bullets include items like placing the lessons in a logical order, using Bloom's taxonomy, and paying attention to diverse learners. Since there is no breakdown, I will assume that each element is worth the same. However, it is not clear that this is the case, nor it is clear what it means to "minimally" address these bullets. And, if following the directions in the list leads to a C, then what must be done to get the full 100 points?

After the lesson sequence, PSTs are also graded on what is called the "Four developed lesson planning overview" (C1, p. 3). It is not entirely clear how this is graded, as it follows the same C grade comment as the sequence. However, it appears that these 15 points, allocated across twelve bullets, will serve as an umbrella grade for all the lessons, even though those lessons are also graded individually. The bullets here seem to apply to general components of good lessons, such as "hook and engage students" and "include assessments that match objectives" (ibid). There is also a comment that is both bolded, in red text, and in title case that says "Substitute Teacher

Ready" (ibid). Thus, it appears that the purpose of these 15 points is to provide Dr. Aldebaran a place to note and comment on general successes or issues that are seen throughout the unit. To emphasize this, it is the only place on the rubric where space is explicitly given to write comments.

Next there are the four lesson sections. Each mini-rubric follows essentially the same pattern (C1, pp. 3-5). There is a "pre-instructional phase" that is dedicated to the objectives, accommodations, and materials. Then there is a "set" which is about motivating, connecting knowledge, and developing a purpose. After, there is an element dedicated to the particular lesson format (e.g. lecture or jigsaw) with sub-elements describing what needs to be included for that particular format. And finally, there is an element about the "closure." Each lesson format has slightly different point values for the elements, as they all have slightly different explanations and variations, but the direct lesson ends up being 60 points, and the other three are 65, so it is roughly even.

Table 5.1 Point Allocation for the Four Detailed Lesson Plans				
Lesson Component	Lecture	Jigsaw	Group Investigation	Cooperative Lesson
Pre-instructional Phase	5	8	8	7
Set	5	5	5	5
Core element	20	25	25	25
Closure	15	15	20	20
Handouts	10	7	7	8
Other	5	5		
Total Points	60	65	65	65

For the Lecture lesson, Dr. Aldebaran writes that the lecture component (the core

element), worth 20 points, must be "clearly outlined in plan" and that "student involvement [is] included" (C1, p. 3). Thus, there is minimal guideline for how to design a lecture lesson, but the rubric does provide some structure. Additionally, this lesson allocates 5 points for how the lesson

will be differentiated to accommodate student needs, and 10 points for the visual presentation and student handouts. Therefore, there is some direction for how to set up this lesson.

For the Jigsaw lesson, the descriptions of the pre-instructional phase, set, and closure are nearly identical to those in the direct lesson (minus minimal changes to the names, such as replace "lecture" with "jigsaw"). It is not clear, then, why the point values change. For guidance on how to attain the 25 Jigsaw points, Dr. Aldebaran provides a short list of sub-components to describe the element. For example, PSTs must include directions and handouts. Packets and handouts are also assessed as their own element in the rubric worth 7 points, separate from the Jigsaw points. it is not made clear how or if these handouts are different. Finally there are 5 points for discussing how the PST will facilitate the social aspect of the Jigsaw lesson. Again, there is not too much directive, but there is some guidance for how to write this lesson.

For the Group Investigation lesson, the points almost mirror the Jigsaw lesson, but there are now extra points in the closure. The reason for this difference is not explicit, but it is possible that it will take more work to list the skills in all the investigation packets than it will to list the skills in the jigsaw expert packets. For guidance on how to attain the 25 Group Investigation points, Dr. Aldebaran provides a short list of sub-components to describe the element. These subcomponents vary in cognitive demand, from "students given choice," to writing directions, to ensuring "active student participation" (C1, p. 5). Again, as with all the subcomponents in this rubric, weighting is not provided. There are an additional 7 points allocated to this lesson for creating "developmentally appropriate" handouts (ibid).

For the fourth lesson, the cooperative lesson, PSTs are afforded a bit of choice. They still must have a pre-instructional phase that is worth 7 points, a set is worth 5 points, and a closure worth 20 points. This time, however, the core element, the Cooperative Lesson, is not so much a

specific lesson type but a category. PSTs are allowed to choose between a "Jigsaw, Group Investigation, Inquiry, or Guided Discovery" (C1, p. 5). Like in the Group Investigation, the development of this Cooperative Lesson is again worth 25 points. This time, however, the elements in the rubric tell the PST that the criteria will vary based on which lesson type they choose, but still give some general comments (they must pay attention to group size, be clear with the lesson plan, and have prepared questions for their students). This lesson also has the Handouts element, which this time is worth 8 points.

Finally, the rubric concludes with 20 points allocated to the bibliography. These points are divided into 10 points for including ten resources, and 10 points for annotating the bibliography. This is the one time in the rubric where it appears to be quite clear how the points will be divided up and how partial credit will be awarded.

Overall, this task description and rubric give a general overview to the PSTs of what must be included in their PPp2 in order to get a C grade, and how points are allocated. It is not always clear why points are divided up the way they are, nor is there an explanation for how partial credit will be awarded. Nevertheless, it does provide an understanding of what Dr. Aldebaran expects her PSTs to complete in order to pass this assessment.

#### Other assignments

Dr. Aldebaran also provided me with information about some of the other graded assignments in her course. These assignments did not come with task descriptions or rubrics, nor did they come with sample work, but I mention them here anyway because it is helpful to know what else contributes to the course grade.

An example of a graded in-class activity is to create a lesson plan that contains a "tiered" activity (personal communication, 5 September 2018). She did not describe what this lesson-type

means to her in our e-mail communication, but she did note that the lesson will need to include not just an activity, but also a lesson title and a list of learning objectives. This is the third time I have seen her mention that a "title" needs to be included.

Dr. Aldebaran also assigns graded online discussions. She says that the topics change from semester to semester, presumably based on the students and their needs. Some topics that have been covered in the past have been "classroom management," "family interactions," or "maximizing learning through differentiation" (personal communication, 5 September 2018). Without the task description, I cannot make any analyses on what must be included or how these discussions are graded. Nor do I have any information on how much weight these discussions hold.

A third way that Dr. Aldebaran grades her PSTs is through reflection papers. PSTs are expected to reflect upon various components of the course. The list of topics that she gave me includes "unit planning," "field experiences" and "learning in the course" (personal communication, 5 September 2018). Again, without the descriptions or rubrics, not much can be analyzed, but it is important to note that PSTs are graded not just on their core assignments, but also in these other ways.

#### Graded PST submissions

Because of when it was in the semester that I interviewed Dr. Aldebaran, and because the PPp2 is turned in as a hard copy, I was only able to closely review and analyze PST submissions of the PPp1. During the interview, Dr. Aldebaran allowed me to quickly skim a sample of a PPp2 from a previous term, but due to time constraints, I was not able to take any notes or make any real observations. I was able to see that a PPp2 was presented in a binder with clear dividers, and

was a hefty submission, but not much else. As I was not allowed to take the sample with me, that is all I can say about that sample.

I have been given, however, two submissions for the PPp1. One is from Electra, who designed a 10<sup>th</sup> grade English unit and the other is from Maia, who designed a 6<sup>th</sup> grade math unit. Electra received a B- and Maia received an A-. Both samples are helpfully annotated with Dr. Aldebaran's comments.

The first thing I noticed when reviewing the two submissions is that Dr. Aldebaran remarks on mechanical, grammatical, and lexical structures. She highlighted areas where Electra missed a space between sentences and crossed out some words in Maia's submission. While it is not clear how much this influences the grade, it is at least noted.

Electra starts her submission with an explanation of her unit. She has chosen a dystopic novel and will use it as a springboard to help students understand issues of violence, forced labor, and hunger in the world. She gives two paragraphs that include the importance of drawing these themes from literature and how they relate to the real world, as well as quotes and statistics about poverty and involuntary labor. At the bottom of the page, Dr. Aldebaran has written her own paragraph seemingly to demonstrate to Electra how this introduction should have been written. In comparison to Electra's introduction, Dr. Aldebaran writes a more concise paragraph highlighting the key injustices in the book and how they mirror situations in modern society. Instead of starting with describing the genre and the importance of using novels strategically in class, Dr. Aldebaran demonstrates that Electra should focus on explaining that the book contains many themes that will afford English students the chance to explore modern-day social injustices. She also leaves a side comment asking Electra how she sees involuntary labor connecting to her students' lives. Electra explains the hunger connection, but not this. Dr. Aldebaran grants Electra a 12/15 on the

rationale, so it appears that despite the rewrite and the missing pieces, Electra was on the right track.

Maia sets up her rationale in a slightly different format. Unlike Electra's block paragraphs, Maia splits hers into three sections. She begins with an introduction on the key math topics that will be covered and with the social setting that she will use to teach these topics. Dr. Aldebaran suggests a rewrite that starts instead with the social setting and then moves into the math. Next, Maia has a section about relevance to the students. She writes several small paragraphs about modern-day hunger and access to healthy foods. Dr. Aldebaran has added no comments, so I assume this means Maia has successfully defended the relevance. Finally, Maia writes a paragraph on why her chosen topic is important to be covered in school. She explains that "statistics and data are used throughout life" and therefore it is important to teach this to students (C4, p. 1). She also talks about the value of teaching topics of social justice in school. Dr. Aldebaran comments that the educational link should be more explicit between her topics and the math. However, Maia does score 15/15 on the rationale, and so despite the areas for improvement, she has demonstrated that she meets the required criteria for excellence.

Next, both PSTs fill out the template. At the top, they both put the grade level, the unit title, and an essential question. Maia has a "topical essential question," to which Dr. Aldebaran adds an "overarching essential question" (C4, p. 2). It is not explicitly clear why Maia needs two types of questions or why Electra does not. Then, both PSTs have a unit goal that Dr. Aldebaran has rewritten. For Electra, the change is to simplify the goal to focus on the interconnectedness between the injustices and actions in the novel (instead of looking at each part individually). For Maia, the change is to make the goal broader. Maia originally included a list of the mathematical skills and objectives to be covered in the unit in her goal, and Dr. Aldebaran's edit focused on the general idea of using data to understand a phenomenon, and suggested that the specifics be moved to a later section. Interestingly, Electra scores 11/15 and Maia scores 13/15. It appears that being too detailed is less problematic than connecting the broad ideas in the wrong way.

In the subsequent box in the template, the PSTs need to list the standards that align with the unit. Electra opts to focus primarily on the English standard around "comprehension and collaboration" (C5, p. 2). Maia, however, uses math standards, social studies standards, and language arts standards. Both PSTs appear to copy the text directly from the standards. Looking at the scores, however, it is a bit confusing. Both PSTs score 5/5 in using the Michigan standards. Electra scores 7/10 on listing out the concepts to be taught in the unit, with the comment that she also needed to reference additional subject areas. Maia gets a 9/10 with no comment. Thus, it appears that there may have been some changes in this rubric from how it is written and how it is used, as mentioned earlier. While it appears that using the standards correctly is out of 5 points, it appears that the grade is actually also tied to the content material.

The third component of PPp1 is the assessment plan. Electra's assessment is a presentation about a "specific social injustice" presented in the novel (C5, p. 3). She then gives a list of topics from which students in her class would be able to choose. Dr. Aldebaran suggests a rewrite to the task to make it more specific. Instead of just considering the injustice, Dr. Aldebaran says that students need to both explain the role of the injustice and discuss how the characters used social action to overcome it. She also adds details about how the final presentation would need to look. Because Electra's description was too broad and missing specifics, she scored 7/10 on describing the assessment. Maia's assessment is a multi-media presentation on food "price and availability in a specific neighborhood in the community" (C4, p. 5). She then goes on to explain how the students are to figure out the pricing and how to present the issues of availability. She also wants her students to "propose solutions" to any food desert issues they might encounter (ibid). The only comment Dr. Aldebaran adds is that Maia needs to make the connection to the problem of hunger more explicit. As a result, Maia scores 9/10 on her assessment.

The next part, listing the content criteria and how they relate to the unit goal, is where both PSTs did the poorest. Electra scored 5/10 and Maia, 6/10. It appears that Electra's problem was to be much too general about what should be included in the final project. While she talks about "research[ing] ...their selected topic" and "report statistics," Dr. Aldebaran remarks that she needs to be thinking about what specific knowledge will be assessed and how this relates to the unit. This is possibly an area of tension because PSTs are expected to have this knowledge before they have taken a disciplinary methods course. Maia's entire section for her criteria has been crossed out by Dr. Aldebaran. Her original submission was just a copied version of what she had put in her section above, detailing the components included in the task and how the students are to go about creating their final project. Dr. Aldebaran, in addition to crossing out her answer, gives suggestions for what should be included instead. She suggests including "what math needs to be in the Product," and a list of the math that will be done explicitly by the students, such as "data presented using Stem-n-Leaf table" (C4, pp. 5-6). However, despite the complete removal of Maia's answer, she still scored higher than Electra. Thus, her crossed out work must have still been on the right track. In the interview, Dr. Aldebaran mentioned that she consciously works to not grade the math PSTs harshly for not teaching a lesson the way she would have, when she was a math teacher.

Following the content criteria, both Maia and Electra describe the product that will need to be produced by their students. Both discuss how their students will work in groups to create a presentation that will be about nine minutes long. Interestingly, when describing the project, they both say essentially the same thing:
Table 5.2 PST Description of Student Group Project								
Electra	Maia							
Students will work together in their assigned	Students will work together in assigned							
teams to create an informative campaign	teams to create a PowerPoint or Prezzi on							
video with an approved topic from the	their neighborhood. The presentation							
social injustice list. The presentation should	should include appropriate photographs,							
include appropriate photographs, video,	video, music, statistics, charts, graphs, and							
music, graphs, and other visual aids. The	other visual aids. The final project should							
final project should be between 6 to 9	be between 6 to 9 minutes long and							
minutes long.	narrated live.							
The video will be evaluated on quality of	The multimedia presentation will be							
understanding of the topic, production	evaluated on understanding of the topic,							
(informative & persuasive), organization,	delivery, organization, and visual appeal of							
and creativity. Group must include a	presentation. Group must include a process							
process paper on how they developed their	paper on how they developed their project.							
project. They are to include an annotated	They are to include an annotated							
bibliography of their research.	bibliography of their research.							
(C5, p. 5)	(C4, p. 6)							

In the table, I have changed the color of the identical words to blue. This suggests that both PSTs built their curriculum off the same model, or else worked together in the development. Both then continue with a bulleted list of what will be assessed in the presentation (without points or weights), with the only real difference being that Electra sorted her bullets into categories. Dr. Aldebaran points out to Maia that a few of her bullets are unclear. However, both Electra and Maia score 10/10 on this section.

Next, both PSTs list the formative assessments that will be used in the unit. Again, their lists are quite similar, with Maia including every one of Electra's assessments on her list, but also adding two more. It appears once again, since they both receive 5/5, that this list is meant to be created using a list that was covered in class or in a text. It is also notable that at this point (although it likely changes in PPp2), that this list does not to be explained or defended, and that just a list of formative assessment strategies needs to be included.

For the last part of the assessment plan, the PSTs present how the students will reflect at

the end of the unit. Once again, their answers are nearly identical.

Та	Table 5.3 PST Description of Student Reflection									
	Electra		Maia							
•	Class discussion: Will take place as an	•	Class discussion: Will take place as an							
	introduction to unit and also as a		introduction to unit and also as a							
	culminating overview of what was learned		culminating overview of what was learned							
	and how students learned it.		and how students learned it.							
•	Student-facilitated formal debrief - students	•	Student-facilitated formal debrief - students							
	will lead the final discussion, which will be		will lead the final discussion, which will be							
	based on visual presentations and questions		based on visual presentations and questions							
	that arise from the summative projects.		that arise from the summative projects.							
•	Teacher-led formal debrief- teacher will	•	Teacher-led formal debrief- teacher will							
	guide students through final culminating		guide students through final culminating							
	notes. (C5, p. 6)		notes.							
		•	Individual journal responses – focus what							
			was learned and how they processed							
			learning. (C5, p. 5)							

The only difference is that Maia includes journals as a method for reflection. By way of

commentary, Dr. Aldebaran suggests to Electra that she include at least one more way to have her students reflect ("Exit Tickets, Journals, or Blogs") and gives suggestions for prompts (C5, p. 6). To Maia, she suggests that she include the journal prompts in this submission. Interestingly, Electra scores 4.8/5, while Maia, whose submission was more thorough, scores 4/5.

It is also interesting that in the bibliography section, both PSTs score 10/10, even though only Maia includes the required ten sources, while Electra includes only nine. It is also interesting that Electra cites an example of a PPp1 that was written by Dr. Aldebaran. It is possible that this is the format that both she and Maia are copying in their submissions.

In summary, here is a breakdown of their scores on the components in the PPp1.

Table 5.4 PPp1 Grade Breakdown for Submissions									
Rubric Element	Possible	Electra	Maia						
Rationale	15	12	15						
Topic for grade level	5	5	5						
Unit goal linked to standard	15	11	13						
Concepts/Knowledge list	10	7	9						
MI Standards	5	5	5						
Description of project	10	7	9						
Content Criteria	10	5	6						
Criteria for Product	10	10	10						
Formative Assessment	5	5	5						
Student reflections	5	4.8	4						
Bibliography	10	10	10						
Total	100	81.8	91						

Maia should have scored a 91 on this assignment (as calculated in the chart), but the given grade was a 92. This difference is likely small, and comes from a mis-add of the italicized section. I have also put in red the areas where one PST scored better than the other. All in all, the major differences seem to come from the description of the course and the listing of the knowledge and standards. Maia's more robust submission earned her the higher grade. Both PSTs, however, scored above the pass requirement for the PPp1.

# Interview with Dr. Aldebaran

In this section, I take excerpts from my interview with Dr. Aldebaran to better understand what, according to her, is and should be assessed in this course. All quotes are from either my notes and/or the transcript and are noted accordingly.

When asked to describe the purpose of her course, Dr. Aldebaran called it "a general methods course" and a "zero to sixty course" (transcript). She explained that this course takes **PSTs** who have likely never written an objective before to be able to understand how to do that up through designing a nearly full unit. As the core assignment in this course is planning the unit (the

**PP** project), the whole course is pretty much designed around helping the **PSTs** to be able to do that. She added that the "weakness is in classroom management" since the focus of the course is more on instruction, and while classroom management is implicitly covered in the course, it does not receive much focus or direct attention (ibid).

Up to this point, PSTs have taken general education requirements, such as about "learning theories" and "motivation," but especially for secondary PSTs, Dr. Aldebaran expects that they have had very little exposure to creating learning objectives or planning lessons (transcript). Her course then covers all the pieces she feels they will need in order to produce a successful PP. She covers "differentiation, writing objectives, writing goals" and other necessary components (ibid).

I asked Dr. Aldebaran about the pre-requisite or co-requisite with Course A, the assessment course. She replied that while PSTs can take the two courses at the same time, she often discourages it. She explained that even when she taught both courses in the same semester, it was challenging to time the components to allow the students to build a curriculum (PP) and assessment (ADP) for the same unit. She added that she has worked with some "very, very motivated" PSTs for whom she did feel comfortable suggesting taking the courses concurrently, but in general, this was not the case (transcript). For everyone else, she recommended that they take her course first, build the full unit, and then take the assessment course at a pace where they could easily write the ADP based on their fleshed out and completed PP. Since the PP does include writing an assessment plan, however, she does talk a bit about assessment in her course. She differentiated this from Course A, and said that she "define[s] formative and summative" but the focus is on the performance assessment in the PP, while the ADP in course A focuses mostly on creating a "final exam" (transcript).

I asked Dr. Aldebaran what she expects her PSTs to be able to know and do by the end of the course. She replied,

I expect that they know how to write a lesson plan, that they understand the difference between student-centered plans and teacher-centered plans and that they know how to differentiate for content, and process, and product. That they know how to identify students in terms of culture and interest and learning styles, preferences, and multiple intelligences so that they can use this to differentiate (transcript).

Thus, she believes that her course prepares PSTs to be able to write quality lesson plans, be able to distinguish between different goals and components, and be comfortable adjusting the plans to meet diverse student needs. In summary, she said she would claim that PSTs who have passed Course C "are ready to now look more in detail at planning within their content area" (ibid). As Course C comes before content methods courses, she provides PSTs with a structure to build future units that the later courses fill in with specific disciplinary knowledge. Dr. Aldebaran added that "it is very nice" to able to focus on the structure separately from the content, and that it helps in the methods courses that the PSTs come in already knowing the basics of lesson planning.

When it comes to planning her course, she says she is "the type of teacher who changes up things all the time" (transcript). She responds to challenges from the previous semester and uses books. The course typically begins with looking the "big picture" and considering the foundations of teaching (ibid). She has students consider their own dispositions, consider their personal philosophies about teaching, and determine where they want to go from here. As she puts it, "you have to know yourself first" (ibid). However, despite this introduction, she does not have them write any formal statements about their dispositions until the end of the semester. From there, the class looks at student and learning preferences. Next, she transitions to considering curriculum maps. For this, she brings in current curriculum maps from all subject areas and the PSTs spend time considering them. As she expects that a curriculum map will be one of the first things they see when they get a job, she feels that starting with it in class important. After, the PSTs look at Bloom's Taxonomy, since Dr. Aldebaran recommends using it to build lesson objectives.

At this point in the semester, Dr. Aldebaran mentioned that she starts modeling the types of lessons she will want to see her PSTs include in their PP. As she teaches the rest of the course, she implements formats like Jigsaws and Cooperative Learning. She covers assessment plans, teacher-centered versus student-centered teaching models, and multiple intelligences. As the semester begins to reach its end, she transitions to a focus on the students and their cultures and how to embed this into lessons. At the same time, PSTs are working on their reflections and peer review.

One thing to note is that throughout this all, Dr. Aldebaran made several mentions of "21<sup>s</sup> Century Skills" (Ananiadou & Claro, 2009). She said that she is a "big fan" of them and therefore the "4Cs are embedded into everything" (transcript). Thus, whatever else she is teaching in her course, she keeps these ideas at the forefront and wants to make sure that her PSTs are considering them in all that they do.

At the very end of the semester, Dr. Aldebaran told me that she gives a reading exam. She described it as "multiple-choice and it's just to encourage them to use the text" (transcript). This exam is open notes and online. While not difficult, she finds that including this exam is a way to make sure the students engage with the printed materials in the course. She also places it at the end of the course because it allows her time to read through all the PPs and give grades, since it takes her about three hours to grade each submission.

I asked Dr. Aldebaran about the choices afforded to her as an instructor for this course. She said that the course itself has a curriculum guide with objectives. All PSTs going through the course must complete the PP. There is also a requirement that all PSTs taking this course write a minimum of three fully developed lessons that include differentiation and assessment. Beyond that, however, how she wants to teach and assess is up to her. She gave an example of one semester when she added a requirement that all PSTs use a field trip to a local museum in their lesson planning. Another semester she required her PSTs to include a lesson on a community problem. She makes the decisions to create these alterations based on a number of factors. Sometimes, she said, it is "because I really see an interest" (transcript). Other changes she makes come from new things she learns from conferences or when she picks up a new technology from a local school. Essentially, she adapts her course to make it meaningful and relevant to her students.

As the focus of my dissertation is assessment, I asked Dr. Aldebaran specifically how she grades and weights assignments in this course. She said that she makes the core assignment, the PP, worth 50% of the grade because since the requirement is to get a C on it to pass the course, she wants it to mathematically work out that not passing the PP will also directly lead to not passing the course. The other 50% of the grade, however, is a mix of the other course components, and she did not give me a breakdown. She did say that the mix consists of classwork, field experiences, journal entries, reflections, professionalism, and participation. She mentioned that she is "really big on reflection" and so there are a number of graded reflections during the course. She also said that she had tried removing professionalism from the grade a while back, and it led to her seeing behaviors that she would rather not see, and thus she felt forced to re-add professionalism to the grade. Similarly, the weight attributed to participation changes based on how the previous semester

went. It was not clear, however, how any of the components, other than the PP, were actually scored or weighted.

Dr. Aldebaran also talked about the logistics of being an instructor for this course. As she has to grade each unit written by all her PSTs (which in the term I interviewed was 48), she requires that her PSTs follow specific templates and present their work in certain ways. She also requires certain types of lesson plans (such as the jigsaw) to be included, as this helps with the consistency of grading.

Interestingly, there were also components of the course that were not assessed. For example, the PSTs are required to peer review each other's PP and the peer review is handed in with the final project. Dr. Aldebaran said that it is a "subtle way to find out how much of [the PP] was last minute, but I don't assess it" (transcript). Thus, Dr. Aldebaran includes checks in her course to help her understand her students that will not influence their grade.

Dr. Aldebaran also referenced academic freedom in her description of the choices she makes for the course. This freedom even extends to how the PP is assigned and assessed. For example, she has split the PP into the two components, PPp1 and PPp2 because she has found that grading the first component first helped lead to better and stronger projects, but that it is not common practice. She said that other instructors occasionally choose to focus on various components of the PP, such as raising the emphasis on writing objectives. Some instructors have their PSTs write full plans for up to ten lessons, while she chooses to only require four full lessons. When she has helped a colleague grade their PSTs' projects, she has used their rubrics, because she felt it was only fair to grade the PSTs based on expectations given to them. However, she says that these differences are small in the big picture, and that it does not really change the claims that can be made about the PSTs once they pass, and if the grade is different, it will only be by a half grade (e.g. B versus B-).

Lastly, I asked Dr. Aldebaran about her conception of fairness when it comes to assessment. Her answer? "Consistency is the biggest thing" (transcript). In order to ensure consistency, she grades all PSTs from the same subject areas together. She also keeps past assignments and grading comments to help ensure continuity from year to year. She spends time researching the subject areas used by her PSTs in their projects so that her personal understanding will not influence the grade. She also said that she works to grade evenly across the disciplinary areas. As her background is in math, she tries to not penalize PSTs if they decide to teach a unit a way that is different from how she would do it.

Overall, Dr. Aldebaran gave the impression of caring about fairness in her course. She aims to "filter it all out" (her biases and opinions) and make sure that she is consistent across PSTs and semesters. She makes changes to the course, uses her academic freedom, and brings in new technology - all with the goal to make the course relevant and meaningful for her PSTs. While it was not always clear how she grades some of the assignments, it was clear that she keeps track of her methods and aims to be a fair and equitable teacher for all her PSTs.

## Looking across

As informative as it is to look at the four indicators separately, much can be learned from considering at them together. First, I was able to build a matrix that matched the core assignments and the bulleted list of goals as stated in the syllabus. I went through each rubric element from the PPp1 and PPp2 and matched it with the bullet that was most closely linked. There was not always a perfect match, but I did the best that I could, and had my colleague, Katie Cook, conduct a reliability check (details for determining reliability can be found in Chapter 3). If it did not fit any of the fifteen bullets, I then decided that the element was either about general mechanics and rule following, or not, in which I labeled it other. I also assigned weights to all the components based on the point values presented in the rubrics. If an element was worth 15 points, I divided it by 500 and found that it was worth 3% of the full PP. If an element contained more than one subcomponent, I divided the weight equally. For example, the element "Planning explicitly articulates how unit will help all students develop understanding" from PPp2 seemed to best fit both "accommodate instruction to diverse learning styles, which are culturally based" and "demonstrate a confidence in all children and their ability to learn," I split the weight across both bullets.

From this chart (Table 5.5), I was able to make a number of conclusions. First of all, bullet 3 was the most assessed bullet in the **PP**. This bullet says, "write clearly stated learning outcomes and apply principles of systematic instructional planning and decision making which lead to student conceptual growth" (C3, p. 1). It is unsurprising that this bullet holds so much weight because as the **PP** is about designing lessons, it makes sense that writing outcomes and planning lessons would be strongly assessed. The next highest weight was general mechanics and rule following. In a project with so many components, it is likely that meeting all the requirements would be assessed. It can also be noted from the chart that many of the bullets are not assessed in this course assignment. However, since the **PP** is only 50% of the **PSTs**' course grade, one can assume that the other 50% will come from the other bullets.

Table 5.5	e 5.5 Matching PP Rubric Elements with Syllabus Objectives																	
			Bullet Assessed															
Rubric Element	% Weight of Full Assignment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Other	Mechanics
Α	3.0%	1																
В	1.0%	1																
С	3.0%			1														
D	2.0%	0.5		0.5														
Е	1.0%			1														
F	2.0%			1														
G	2.0%			1														
Н	2.0%				1													
I	1.0%				1													
J	1.0%							1										
К	2.0%																1	
L	0.0%																	1
М	0.4%																	1
Ν	0.6%																	1
0	2.9%												1					
Р	2.9%			1														
Q	2.9%																	1
R	2.9%									1								
S	2.9%					0.3 3	0.3 3					0.3 3						
Т	2.9%				1													
U	2.9%																1	
v	0.25%							0.5									0.5	
w	0.25%					0.3 3		0.3 3		0.3 3								

Table 5.5 (Continued)																		
										В	ullet A	ssessed	1					
Rubric Element	% Weight of Full Assignment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Other	Mechanics
Х	0.25%							1										
Y	0.25%												1					
Z	0.25%					1												
AA	0.25%												1					
BB	0.25%																1	
CC	0.25%																	1
DD	0.25%																1	
EE	0.25%																1	
FF	0.25%			0.3 3		0.3 3												0.33
GG	0.25%					0.3 3	0.3 3		0.3 3									
нн	1.0%			0.3 3		0.3 3												0.33
II	1.0%					1												
JJ	1.0%							1										
КК	4.0%							1										
LL	2.0%																	1
ММ	3.0%	0.5		0.5														
NN	1.6%			0.3 3		0.3 3												0.3333
00	1.4%	0.5																0.5
РР	1.0%							1										
QQ	5.0%				1													
RR	1.0%							0.7 5					0.2 5					
SS	3.0%			1														

Table 5.5 (Continued)																		
			Bullet Assessed															
Rubric Element	% Weight of Full Assignment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Other	Mechanics
TT	1.6%			0.3 3		0.3 3												0.33
UU	1.0%							1										
vv	1.4%							0.5										0.50
ww	5.0%				0.5													0.50
XX	4.0%	0.5		0.5														
YY	1.4%			0.3 3			0.3 3											0.33
ZZ	1.0%							1										
AAA	1.6%																	1
BBB	5.0%																1	
ссс	4.0%			1														
DDD	2.0%																	1
EEE	2.0%																	1
Total	Weight (%)	9	0	24	13	4	1	11	0	3	0	1	4	0	0	0	11	18

I next grouped the bullets to see what trends I could assess. I first started with the five themes set forth in the syllabus, while still keeping other and mechanics.

Table 5.6 PP Element Weight by Syllabus Themes							
Theme 1: Student learning	20%						
Theme 2: PST Knowledge	44%						
Theme 3: Reflection	0%						
Theme 4: Professional Dispositions							
and Communication	4%						
Theme 5: Diversity and Democracy	1%						
Other	11%						
General Mechanics and Rule							
Following	18%						

From here, I could see that the PP weighted PST knowledge most highly, and despite Dr.

Aldebaran's comment in the interview that she cares strongly about reflection, it was not assessed in the core assignment.

I also grouped the bullets according to the themes that I found when I analyzed the bullets.

Table 5.7 PP Element Weight by My Themes						
Designing Curriculum	52%					
Modes of Communication	4%					
Considering Students	14%					
Transitioning from Student to Teacher	0%					
Engaging in Controversial Topics	1%					
Other	11%					
Mechanics	18%					

From this, I found that the bulk of the weight from the **PP** was centered on designing curriculum. This made sense because the **PP** is a curriculum-planning project. Excluding "other," the next most amount of weight came from considering students, which also made sense seeing how strongly the syllabus stressed being a caring teacher and considering the backgrounds and diversity of the future students. Since I did not have rubrics for the non-core assignments in the course, I also aimed to

determine, based on my conversations with the professor, how all the bullets might be assessed in

the course. This is what I found:

Table 5.8 Matching Syllabus Objectives with How Assessed	d in the Course
Bullet (C3, pp. 1-2)	How assessed
Select and analyze curricular content and goals as well as adaptations to such within the diverse caring community of the school	<ul> <li>Class assignment of reading a curriculum map</li> <li>PP</li> </ul>
Critical discuss various philosophies and trends in curriculum	<ul><li>Class discussion</li><li>Teaching philosophy paper</li></ul>
Write clearly stated learning outcomes and apply principles of systematic instructional planning and decision making which lead to student conceptual growth	• PP
Design and use varied lesson and learning strategies and analyze their effects	• PP
Accommodate instruction to diverse learning styles, which are culturally based	• PP
Explain special considerations educators should address when planning teaching and learning experiences with Native American, African, Hispanic, and Asian American children.	• PP
Reflectively describe ways which increase student attention, motivation and learning given the school environment in general and the classroom in particular	<ul><li>Journal entries</li><li>Reflection papers</li><li>Group presentations</li></ul>
Show respect for all children and the cultures they bring to school	<ul><li> Reflections</li><li> PP</li></ul>
Demonstrate a confidence in all children and their ability to learn	<ul><li> Reflections</li><li> PP</li></ul>
Make the transition from scholar/student to teacher-student	Professionalism
Show willingness to address controversial issues	<ul><li>Class discussions</li><li>Journal entries</li><li>Reflection papers</li></ul>
Demonstrate success in using varied communication strategies, including questioning and discussion	• PP
Use technology to learn and communicate	<ul><li>Group inquiry project</li><li>Course presentations</li><li>Online discussions</li></ul>
Describe ways to establish positive parent-teacher communication	• Journal
Demonstrate positive interactional skills during small group discussion and tasks	Class work

One thing I noticed while making this table was that the reading exam did not fit well into any of the bullets. Also, even though I was able to guess about how the bullets were assessed, I still could not determine weights or scoring. Nevertheless, I was able to find places where each bullet was assessed in some way.

## Summary

Course C is a complex course that guides PSTs through their learning about curriculum and unit development. Dr. Aldebaran designs her course with the core assignment in mind, and plans lessons and activities that prepare the PSTs for success on the PP. In addition to focusing purely on curriculum structure, Dr. Aldebaran also focuses on the type of teacher that the PSTs will become. The core assignment and the other components of the course grade come together to assess how well the PSTs have mastered the beginning stages of curriculum development, as well as on their ability to reflect upon their growth.

## Chapter 6: Course A

"Some things are indeed hard to assess, I'll give you that, which means sometimes we don't do it, even though we should." -Dr. Polaris, interview

# Overview

Course A is an assessment course for PSTs in the middle of their program. This course takes place after the general education courses, but before subject-area methods courses and student teaching. Course A is taught by Dr. Polaris most of the time, and currently, he teaches all the sections. In the semester when I collected data, he was teaching three sections of this course. Two of the courses met twice a week, and one course met once per week.

For continuity, all documents related to this course are marked by the letter A and a number. All people associated with this course have names attached to the Big Dipper constellation, with Dr. Polaris (the North Star) being the central focus, as he is the professor. **Syllabus** 

The syllabus of Dr. Polaris' course begins with a list of the "essential outcomes" of the course (A9, p. 1). These are enumerated and bolded, which in addition to being at the top of the syllabus, helps to emphasize their importance in the course. The terms are worded using student-centered language, using a "Students will..." format that matches how PSTs are often taught to create objectives for their own lesson plans (ibid). Thus, this format here parallels this and even models it (perhaps subconsciously) for the PSTs. All the outcomes are then listed as verbs, which indicate that these outcomes are not just a passive list of knowledge to be acquired, but behaviors that the PST will need to be able to do. This then indicates that they will be expected to perform in some way.

The first essential outcome talks about the principles of assessment. Interestingly, the verb used for this outcome is "explain" (A9, p. 1). PSTs, after taking this course, are not expected to just know or understand the principles, but to be able to explain them. This suggests that when this outcome is assessed, even if assessed on a written exam, PSTs will need to be able to describe how the principles work. This also suggests an ownership of the knowledge. The second essential outcome is about being able to "meaningfully critique" assessments (ibid). There is a bit of subjectivity in this outcome. Does "meaningful" mean to "correctly" critique, or is any critique that is supported with evidence justifiable? The third outcome is to "construct quality tests" (ibid). As an introduction to assessment, this outcome is expected. It also seems to build off the previous two outcomes since to construct quality tests, one needs to know what quality looks like and anticipate potential critiques. The fourth outcome is to analyze and use assessment data. Because assessment scores are only useful if the scores mean something to the assessor, it makes sense that in an assessment course, PSTs would learn how to take the results from an examination and use those results to make decisions. The last outcome is to "advance... levels of professionalism" (ibid). This outcome stands out from the other four because it is more about the person taking the course and less about the knowledge to be developed through taking the course. While the other four outcomes relate to how well the PST can understand, develop, and analyze assessments, this outcome focuses on the behavior of the PST while learning and growing as a future teacher. It also assumes that each PST will come in with a current level of professionalism, which will then be enhanced through taking this course. Thus, while four of the outcomes relate to assessment, one is focused on the person.

The syllabus then lists fifteen topics that will be covered during the course. These topics include understanding intelligence, basics of evaluation, different test forms and rubrics, and

reporting. For a course that only meets for fifteen weeks, and with an exam on two of the weeks, it appears that this material will be covered quickly.

The next half page is dedicated to policy of the course, which is reinforced by another half page about professionalism on page 3. Despite this essential outcome only being stated as 10% of the grade, it definitely gets a lot of attention in the syllabus. After the initial mention, an additional 270 words are dedicated to explaining professional behavior in the course and the full syllabus is only 870 words. (For comparison, outcome one has 72 words, and outcome two, three, and four have no additional mention). Procedures for dealing with tardiness are described in detail and methods for making up absences and personal responsibility are laid out. At least from a syllabus point of view, it appears that personal behavior in this classroom matters. Dr. Polaris sets a tone of professionalism and describes what it means to him and what he expects of his PSTs. Thus, despite the 10% number, it appears that this component is crucial for passing this course.

As part of the professionalism just mentioned, tardiness appears to matter strongly to Dr. Polaris, as it is given its own paragraph. As stated, "excessive tardiness" can lead to a 0% in the professionalism grade (A9, p. 3). However, one of the sections meets twice a week and the other section meets for a longer period of time just once a week. It is not clear, then, if "excessive tardiness" is the same for both classes. Does Dr. Polaris care about what percent of times one comes in late? Is it a raw number? Is it how many minutes are missed? There seems to be a potential for bias because a student who is late to class twice in one section might have a different feel to being late twice in the other. The subjectivity of this professionalism grade is acknowledged explicitly in the syllabus. The difference between a 90% and 100% in professionalism depends on the "judgment" of the professor (ibid). This subjectivity continues in the explanation of the policy, with words like "strong" used to describe the necessary engagement in class and in groups (ibid). Next is a description of all graded components in the course. There are two exams, a midterm and a final, that are worth collectively 25% of the course grade. Dr. Polaris does provide study guides, and mentions here that the bulk of the exam material will come from class lectures and activities, and not course readings. This emphasizes the importance of regular attendance in class and active participation. This is not a course where doing the readings at home and showing up for exams will lead to a passing grade. There are also two major assignments. These are not described in the syllabus other than to be given course percentages, and to mention that while the core assignment is worth 50% of the grade, since it is the required assignment for the department, one needs to score 70% on this assignment alone to be allowed to pass (regardless of the overall grade). This policy appears to be consistent across all required courses for the teaching major at this university. Dr. Polaris also says that there will be non-graded assignments that are mandatory to complete. Even though ungraded, they must be completed in order to receive a final grade in this course.

In the remaining syllabus pages, there is a conversion chart from percentages to letter grades, and a brief schedule of course dates. For the grades, Dr. Polaris states that there will be no rounding. For the schedule, Dr. Polaris includes dates for the midterm and final, as well as due dates for the major assignments. It appears that for sections that do not meet on the same day, due dates and exams are due as close as possible for each section.

Thus, what is important to the professor for passing this course according to the syllabus? It appears that PSTs need to meet all five outcomes through actions, not just understanding; they need especially to maintain high levels of professionalism as defined by Dr. Polaris (more than the 10% seemingly allocated); they need to complete all assignments, even if they are not for a specific

grade; the core assignment is critical for passing this course; and PSTs will leave this course with a foundation on the nature of assessment.

### Major course assignment descriptions and their rubrics

In this next section, I review the task descriptions and corresponding rubrics for the two major assignments in this course.

#### Analysis project (AP).

This assignment is worth 15% of the course grade; therefore, it holds significant weight in the course and, while not the core department assignment (although from the interview, it seems to be a constant assignment regardless of the professor or term), it is assigned by the professor each semester he teaches the course. The AP is done in four stages, Planning, Administering, Analyzing, and Reflecting.

In the Planning stage, the PST works with a classroom teacher (CT) to decide upon an examination that is coming soon in class. There is some freedom in the choice of the exam, but it must be one that the CT is willing to allow the PST to administer. Dr. Polaris provides the PST with a blanket letter to give to the CT, explaining the purpose of this task.

In the Administering stage, the PST is present in the classroom on the day that the exam is given and is in charge of administering the examination. This includes explaining the directions and answering any student questions (if applicable) during the exam. Additionally, the PST needs to take note of the student behaviors and reactions before, during, and just after the exam. If there is a template to help guide the PST, it is not included in the task description.

In the Analysis stage, the PST, under the guidance of the CT, is to grade the student exams and create a test sheet. While this is not explained in the task itself, based on seeing sample assignments and talking with Dr. Polaris, it is clear that this sheet is following a method that is discussed in class. Once the sheet is made, the PSTs are required to analyze the results and look for possible reasons why students may be getting certain questions wrong.

In the Reflection stage, the PST needs to consider how the experience of the AP has informed their development as a teacher. There are guiding questions that can be answered, but it appears that the goal is to get the PST thinking in general about administering an examination.

The rubric for this assignment tells the **PST** what they need to do to get a 100%, but there is no given weighting scale or intermediary scoring information. The elements are:

- 1. You provide a clear and detailed explanation of classroom context.
- 2. You provide a detailed description and critique of administration.
- 3. You provide a clear, accurate and precise presentation, explanation and interpretation of results.
- 4. You use proper tools in preparing data summaries.
- 5. You provide appropriate and clear suggestions for use of results.
- 6. You provide a quality reflection.
- 7. You follow the guidelines provided and turn in an easy-to-follow report, cleanly presented, on time, mostly free of mechanical errors (A2, p. 3).

The seven elements are listed as necessary for the 100%, followed by this comment: "As these requirements are not met, the grade will be lowered accordingly" (A2, p. 3). Without the commentary that likely accompanied this task description in class, it is unclear how to understand this comment. Are all seven components equally important? Additionally, the vocabulary used in the descriptions of each component is subjective, and without the additional commentary, can be hard to decipher. Dr. Polaris uses terms like "clear and detailed" or "appropriate" or "quality," which may not invoke the same ideas in everyone. This subjectivity leaves room for possible

misinterpretation, but also gives the instructor significant freedom to understand the goals of each **PST** and not be boxed into a specific point score. Overall, when reading just this task description, it is hard to see what is most important. What is clear, though, is that following the instructions and including all components is at least a minimum requirement for acing this assignment. This theme of following instructions continues to be present here, as it was in the syllabus.

## Assessment development project (ADP).

The ADP is the core assignment of this course. As with all core assignments in this College of Education, a score of 70% or better on this one assignment is required for passing this course, in addition to passing the course itself. The general idea of this assignment is uniform across all sections of this course, regardless of the term or the instructor, as it is a departmental assignment. The exact components and rubric, however, are still under the jurisdiction of the individual course instructors.

According to the task description, this assignment is designed to measure three of the five essential outcomes: demonstrating knowledge of testing principles, constructing assessments, and professionalism. To do this, the PST is required to construct a "summative assessment plan" (A1, p. 1).

Like the AP, the ADP is composed of four stages, or as named in the task description, "parts" (A1, p. 1). This time, however, the four Parts are given their own weight, which provides some insight into the relative importance of each. The four Parts are described next in detail. First, though, there are several big picture aspects of the ADP that should be discussed. The second page of the task description is allocated to giving general guidelines for the task. Most of this page is dedicated to structural directions, such as what the title page should look like, how the assignment should be formatted, and reminders about language and spellcheck. The additional instructions remind students about the rules on plagiarism and the scoring requirement. It is not clear from the description if this adherence to detail is for ease of grading or because it has some inherent purpose within test development. In either case, it appears that these rules must be followed and it points back to the importance of "professionalism" in this course.

Throughout the task description, clear instructions are given. Sub-components for each Part are numbered, key elements are bolded, suggested word lengths are given, and as summary, a rubric for what 100% on each Part would require is provided. Like the AP, rubrics in the ADP do not give relative weights to the sub-components, and descriptions use subjective terminology.

Part 1 is worth 20% of the grade and is allocated toward describing the unit for which this assessment would be used, and creating an assessment blueprint. First, PSTs need to determine what hypothetical unit they want to be assessing. If PSTs are currently enrolled in the curriculum course, they may use that unit, but it is not required. PSTs are given the freedom to choose whatever unit they want (within reason, and related to their major), and the choice of unit is not really graded. Because this course is taken prior to a disciplinary methods course, it makes sense that Dr. Polaris does not grade PSTs on the appropriateness of their unit. As long as a generally reasonable unit is described, has an introduction that is "clear and helpful," and backed with Standards, then it is sufficient (A1, p. 3). Second, PSTs need to build a test blueprint. In the blueprint, PSTs need to incorporate outcomes for the students, three "cognitive complexity columns," assessment type, and relative weights for all (ibid). This will end up looking like a matrix. Third, students need to write a paragraph or two defending their choice of weights. This explanation is to be "general," which may be because students are not yet expected to fully know their content yet, and making a strong blueprint with real weights would require this knowledge (ibid). Fourth, students need to describe and explain their choice for the length of the assessment. In the second part, PSTs actually use their blueprint and hypothetical unit to create a traditional assessment. This component is worth 40% of the grade. Dr. Polaris provides a list of item types (and number of items for each type) for the PSTs to choose from. The completed assessment must be formatted as it would be formatted for students. Then, PSTs need to create an answer key for their assessment and provide rubrics for any essay tasks. Next, PSTs need to create an additional item bank of supplemental items. These must be formatted like a test and include directions, but an answer key need not be provided. Lastly, the PSTs need to imagine that one of their students has a special need and must describe how the assessment would be modified to meet this need. Inclusion of all these pieces, plus adhering to general assignment guidelines and being neat and clear, is needed to receive 100% on this component.

In the third part, which is worth 30%, PSTs need to reimagine their assessment and design an alternative assessment to assess the same unit. First, PSTs must describe the assessment, including details about cognitive demand and purpose. Second, PSTs must create a task description sheet and format it exactly as it would be given to students. This must include the purpose, the procedures, a time frame, a rubric, and a description of the final project. Additionally, this description must anticipate student concerns and respond accordingly. While this addendum is not described in detail here, it likely refers to concerns like choosing partners or how to help students develop additional background knowledge, as this is how Kochab completes this component in his submission. Interestingly, in the summary rubric for this section, attending to student concerns is not listed as part of the necessary components for receiving 100%. It is possible that this is wrapped into the description sheet component.

Finally, the fourth part is a reflection worth 10% of the grade. There are two required reflections. In the first, **PST**s are asked to reflect upon the full experience and describe how the

overall process "measured [their] classroom assessment knowledge and skills" (A1, p. 7). This is a meta-reflection where they are asked to consider if the ADP is a good tool for the course. For the second reflection, PSTs are given a list of several goals set forth by the Teacher Education Department at their university and the PSTs are expected to explain how this assignment did or did not help them reach these goals. These goals are:

- 1. Demonstrated knowledge of state and national standards
- 2. Learned how to establishes {sic} high level learning goals for your K-12 students
- 3. Learned how to use traditional and alternative forms of assessment
- 4. Learned how to adapt instruction and assessment for a diverse population
- 5. Learned how to use technology and have become technologically literate
- 6. Reflected on teaching and on K-12 student success (A1, p. 7).

Between both reflections, PSTs are expected to write two to four pages. In the rubric, PSTs are reminded to be thoughtful and to provide a critical analysis, as well as to reflect upon all the components and to use examples.

# Thinking about the AP and the ADP together.

Taking the two major assignments together, which comprise 65% of the total course grade, we can understand a bit more about what is important in this course. First, the attention to rule following and professionalism is highlighted. While the **ADP** makes the professionalism link explicitly, both put considerable emphasis on the structure of the assignment. **PSTs** are expected to follow instructions, stick to the guidelines, and complete all components. While the professionalism grade may be a separate course grade, it is not truly separate here. In fact, much of the grade from both of these assessments adheres to it. Second, both assignments have rubrics to describe how the assignment will be graded, but they are not rubrics in the traditional sense (with a matrix and score levels). Instead, Dr. Polaris' rubrics are a list of components that must be included in order to get 100%. Using the wording of "will earn a score of 100% if" suggests that 100% is the starting point and that points will be deducted as necessary. This differs from how Dr. Aldebaran grades Course C, where her rubric lists what is needed to get a C, and then PSTs can get a higher grade if the expectations are exceeded. Thus, the emphasis on the target grade is shifted depending on the model. With Dr. Polaris' model, it seems to appear that what is expected of PSTs is 100%. It is not clear, then, how likely it is for someone to get a low grade on this assignment if they include all the components.

Third, there seems to be a clear focus on structure, rather than substance. PSTs are not graded on their content of their examinations, nor on the accuracy of their claims about students and the students' learning of the particular units, as shown by the accommodations not being graded on accuracy. In both assessments, the focus appears to be on the PSTs' ability to demonstrate an adherence to a development and analysis format, even if the actual analysis is faulty. For example, PSTs need to provide a reasonable unit in the ADP, but it will not be assessed for appropriateness for future students. Similarly, PSTs need to suggest reasons for why students are scoring poorly on a particular test question for the AP, but they will not be graded on the correctness of these hypotheses.

## Graded PST submissions

As useful as it is to look at a syllabus and the task descriptions and rubrics to understand how PSTs are assessed in a course, what makes it real is also looking at graded assignments. I was fortunate that Dr. Polaris gave me three graded assignments, two AP assignments and one ADP. The AP samples include markings and comments by Dr. Polaris, while the ADP sample is a clean

copy with no notations. All three were of **PSTs** who scored well, which afforded me the opportunity to see what excellent work looked like, although not the opportunity to see the pass line. In this section, I will summarize what I was able to learn about the assignments from these samples. I read through each with lenses focused on what was being assessed, what mattered, and what could be "wrong" without penalty.

## Graded AP.

The two PSTs for whom I have sample work on the AP, Yildin and Anwar, scored within one point of each other. Yildin scored a 97% and Anwar a 96%. It appears that Yildin lost his points in his introduction of the examination he is analyzing. According to the task description, one is to "Clearly state the purpose of the assessment. List outcomes being assessed." Yildin merely states that the purpose of the examination is to "ensure that all students met the ... benchmarks for ... semester one" (A6, p. 2). Anwar, on the other hand, in addition to stating that the purpose was to "evaluate student learning of the material in chapter 6," lists the knowledge being assessed on the examination, and receives the comment "nice clear intro" (A7, p. 2). When the task description says that the PSTs need to state the purpose of the examination, merely stating that it was to evaluate student knowledge is a sufficient explanation, and PSTs do not need to look deeper. When I first read Yildin's sample and comments, I thought that the fault was to not go into more detail about the purpose for this individual exam, as he had not talked about how it fit into the term and why the test was designed the way it was. However, seeing that Anwar received full marks, I can now see that Yildin's error was to not also list the content objectives. Neither student gives a deep purpose. Thus, I can conclude that Dr. Polaris wants PSTs to understand that examinations are given to assess specific knowledge.

It appears that Anwar lost her points in her description of the administration of the assessment in the classroom. This perplexes me because this information (about class size and classroom set-up) is not included at all in Yildin's assignment. This leads me to believe that it was not that her description of the classroom was lacking, but that she was supposed to instead describe the lead up to the examination in more detail. Yildin describes how he and his CT worked together to develop the examination and prepare for this project, and Anwar only mentions that she prepped by watching her CT administer the exam during earlier periods. Thus, while the comment says that Anwar could "say more," it appears that it is not just the lack of words, but specific content that is missing (A7, p. 3). This then indicates that the write-up of the AP was meant to reflect upon a more detailed interaction between the CT and the PST. In the task description, PSTs are told to "discuss the project with [their] cooperating teacher" (A2, p. 1). Based on the comments in the submissions, this discussion was intended to both prepare for the administration and be noted in the report.

While not directly assessed, it appears that the relationship between the CT and the PST can influence the performance on the AP. Yildin makes frequent references to what he has learned from his CT, and it results in strong descriptions of the preparation, the administration, and the reflection. Anwar, on the other hand, seems to be less connected with her CT, and while she does note some collaboration, she describes her experience with a more helper-type description, talking about how she helped her CT catch a student who was cheating. This difference is also notable when Yildin talks about how he uses his CT's strategies for getting classroom attention, while Anwar talks about the struggles to get the attention in her class. It is possible that these differences stem not from the relationship, but from the personalities of all involved. Nevertheless, it indicates that success on the AP might be partially influenced by both

personality factors and placement. PSTs in placements with CTs who are more reserved and less likely to share knowledge (or PSTs who are shy and less likely to communicate with their CT) will likely struggle more on this project that PSTs who have strong relationships with their CT, and could even possibly learn less from the assignment.

Both Yildin and Anwar mention that their examinations are multiple-choice. While it is not required in the task description to use a multiple-choice exam for the analysis, I wonder if it makes succeeding on this project easier. If nothing else, it likely makes the scoring of the examinations faster, which will allow for more time for analysis before the assignment is due. Without seeing samples from PSTs who struggled with this project, however, I cannot claim that this will make a significant difference. I can only point out that it might be something to consider in future research.

Part of administering the examination requires the PST to observe the student behavior. It interesting that Yildin and Anwar come to very different conclusions from similar behavior. Yildin mentions that several students during the examination seemed to relax and that this was "deceiving" because these students ended up with low scores (A6, p. 3). Anwar, on the other hand, mentions that "it [is] pretty clear which students [know] the material" (A7, p. 4). I noticed that while the task description requires the PST to take notes on student behavior and reactions, there is not a template provided nor is there a reference to observing in a similar manner as done in class. Therefore, I am left to conclude (perhaps incorrectly) that the PSTs are given the freedom to observe and note in the way that feels most comfortable to them. Yildin walks around with a notebook and Anwar sits at the front of the room. Both PSTs describe the behaviors that they notice, but they are quite different and described differently (as shown above). It appears that it is expected that PSTs know how to observe student behavior and that it is a skill that is expected for this course but not taught.

Both Yildin and Anwar provide a type of scatterplot to display the scoring distribution of the examination that they administered and scored. Both have positive comments next to it from Dr. Polaris, indicating that this is what expected to receive full points for this component. However, Yildin has an interesting situation where his y-axis (number of students) has the number 2 written twice, and with no point actually landing on the lines that correspond with the labels. Anwar does not have a y-axis at all. This indicates that the purpose of this component is to display the work, but that the accuracy of this display does not factor into the grade. What does matter, however, is that the data is displayed in this format and that there are also given the mean, median, and mode.

Another interesting thing about the work is that Anwar writes how many tests she graded, while Yildin does not. Anwar also gives reasons for why every single question on her examination may have been gotten wrong, but Yildin only gives reasons for 11 questions. Part of this discrepancy might be that Yildin's test had 84 questions and Anwar's had only 20. However, because the task description does not say how many need to be analyzed and both these students got full points here, it appears that what only matters is that analysis is done, and not quantity.

What also does not seem to be evaluated is how well these questions are analyzed. The PSTs are expected to give reasons for questions that many students got wrong, but they are not expected to look at the student data to determine why. Yildin often says things along the lines of the students "likely needed to spend more time studying" (A6, p. 6), and Anwar refers to why she found the questions challenging when she attempted the test (A7, p. 9). From my own analysis and review of the PSTs' work, I also noticed that the grading did not depend on the correctness of these analyses. Yildin even managed to analyze the wrong question compared to what he claimed (he suggests that question 30 was challenging because the background knowledge mentioned

climate, but question 30 is about football), but Dr. Polaris does not deduct points for it. Thus, the AP is checking that PSTs are looking at wrong questions and coming up with hypotheses, but is not concerned with the accuracy of these claims.

However, both PSTs clearly spent time analyzing the data and looking for reasons. Anwar took the test as a student to see where she found the test to be most challenging. As she took the exam, it appears that she tried to find ways to get the wrong answer, and then if multiple students got a question wrong, she attributed it to this error. For example, she found that she mistook  $5^{\circ}$  as 5<sup>\*</sup>6 on one question, and assumed that the students might do the same (A7, p. 9). Yildin, in the question mentioned previously, spent time looking at the wording of the questions and looking for potential bias (e.g. with using climate as the context). Both PSTs also looked beyond the test questions and wanted to find external reasons for errors. Yildin researched one student who did particularly poorly and checked with the CT to confirm his conjectures (A6, p. 6). This student had put his head down during the test, and Yildin found out about home life issues that were interfering (potentially) with in-school behavior. Similarly, Anwar checked with her CT and found that many students had either not been completing their homework or had transferred from another teacher, and that this might be leading to lower test scores. Anwar did not confirm her analysis by matching the students with low scores to these students, but she did show some initial research into her students' backgrounds. This indicates that part of what made both PSTs' projects strong was that they looked for understanding both within the exam itself and with the classroom behavior. Demonstrating both skills appears to be important for success on the AP, even if the task description and rubric do not say this explicitly.

Interestingly enough, while the AP task description does not appear to be measuring or developing classroom presence, both samples mention that this project helped with this skill.

Yildin talks about using the CT's method for capturing attention (A6, p. 2) and about learning to answer student questions without giving away the answer (p. 3). In his reflection, he even states that as a result of this task, "I have a stronger sense of self-confidence in the classroom" (p. 23). Anwar also stated that this task was "a great opportunity to practice being in front of students" (A7, p. 10). While not an assessed portion of the AP, it does nevertheless appear to be an element of it.

Finally, when looking at both reflections, it appears that the PSTs have learned to demonstrate what they know and feel comfortable discussing their success. I cannot tell if the task itself was wonderful, or if the students know that being positive in a turned-in assignment is a good strategy. Yildin calls the task "beneficial" (A6, p. 23) and Anwar calls it "very valuable" (A7, p. 10). This may be written because it feels expected, but nevertheless, if nothing else, it is clear that the PSTs have claimed the benefits of the AP.

In summary, from looking at two samples from PSTs who scored high on the AP, I have learned that a number of things matter in this assignment. First, following directions matters. Yildin lost points for not listing specific objectives of the examination, and this seemed to be the purpose of describing the "purpose" of the examination. Directions also seemed to matter when Anwar described her classroom set-up, when it appeared that the implied goal was to talk about the preexam set-up. Thus, giving detailed descriptions was not enough – rules needed to be followed. Second, there appeared to be an implicit value that was attached to the relationship between the PST and the CT. Yildin, who seemed closer to his CT, was both more involved in the examination process and was able to write more about it, and he received more positive comments when he explained how he prepared with his CT for the administration. Third, while observation of student behavior is a requirement for this assessment, how this ought to be done does not appear to be covered in this course. It appears to be a pre-requisite that is expected. Fourth, there was a specific model for how to set up student test score analysis that was expected to be followed. Both Yildin and Anwar were praised for their creation of the scatterplot. Fifth, existence of analysis is more important than accuracy of analysis. For this task, what seemed to matter was that the PSTs looked at the data and tried to make conclusions, but the veracity of their conclusions did not factor in to the grade. Sixth, learning about classroom management appeared to be a side effect of the AP. While not the stated purpose, or even assessed, both PSTs mentioned growing in this area as a result of completing this task. Seventh and lastly, this task rewards positive reflection upon the process.

In addition to the seven points above, there were a few additional things I learned from looking across the two documents. This assignment strongly indicates that the course is a midprogram course. The purpose of this task is to gain experience, and to practice giving, grading, and analyzing an assessment. Using the data analyses as an example, practicing using a given structure seems to be the goal. The implication is that getting used to the process now should translate to successful analysis later once the PSTs have taken more content-focused methods courses and gained more classroom experience. Poorly defending correlations or making potentially false conclusions is not currently a concern. Another note is that, without seeing multiple samples, I cannot know if all PSTs choose multiple-choice examinations to analyze, and if this choice makes the AP easier or harder to do. It appears that the analysis format is straightforward for a multiplechoice test, but there might also be a format provided for other test formats. Another concern is that I am not able to ascertain if this task is easier or harder with certain grade levels and subject areas. It is possible that this could change the difficulty level of this task. Lastly, I wish I could see a pass-line sample, as I can learn much from seeing a sample that is excellent (because it allows me to see what it looks like to enact the full purpose of the task), being able to see a pass-line could

help me better understand what is weighted the highest and what can help a PST to just get the necessary grade for completion. I recommend that future research analyze pass-line submissions.

## Graded ADP.

The sample work I have from the ADP was done by Kochab. This sample is a bit different from the ones for the AP because it is a copy of the original, and does not contain any of Dr. Polaris's markings. It also does not contain an exact grade, although Dr. Polaris informed me that this work scored in the 95-100% range. Thus, when I am analyzing the task I am going to assume that whatever is present is indicative of what Dr. Polaris wished to see when he handed out the assignment.

One of the things I had to constantly remind myself when analyzing this submission is that this task was not assessing a PST's content knowledge. As this course comes prior to any disciplinary methods courses, the focus of this task is on structure development. Thus, as the students are to create a fake unit to put a final assessment at the conclusion, they are not going to be assessed on the plausibility of this unit actually working in a classroom setting, nor are they assessed on how well a test would measure this knowledge. These two plausibilities require knowledge that the PSTs have yet to learn. Thus, I had to remember to not read with the lens I would use to assess a unit and final exam of a current teacher. I had to suspend my disbelief when Kochab's unit included reading two full Shakespeare plays in one high school English unit.

Another notable thing about this submission is that it is 33 pages long, including the title page. While it is double spaced and one-sided, the **ADP** is still a significant undertaking. With all the components and the sheer length, from the outset it looks like a project that will take considerable time to complete, and thus feels appropriate for a core assignment that will influence the passing of this course. For Part 1 in the ADP, Kochab, as instructed, gives a general description of the unit he is assessing and spends a page detailing the outline of the unit prior to the assessment. When asked to "discuss any prior knowledge or skills" (A2, p. 3), Kochab mentions that in order to read the two plays, his students will need to know how to read Shakespearean English and will want to have some background on the "cultural and political situation in England at the time these two plays were written and performed" and therefore he will start his unit with lessons on these topics (A3, p. 3). I find it interesting that the prior skills comment is taken to mean that he needs to mention what he needs to teach in the unit. In other contexts, I have heard of prior skills as describing what you expect the students to already know when entering the course. Based on this sample, however, my understanding was not the intended one, or at least not enough of a difference to significantly influence Kochab's grade.

Part 1 also includes a blueprint, and Kochab's has clearly followed the required format. He has six student objectives, and they are all written with the "Students will be able to" stem. He separates out his objectives for the traditional test and the alternative assessment. He includes three cognitive levels, and he places his test components accurately between his objectives and his levels. For example, he places a vocabulary matching section as a low cognitive demand task to assess the objective about being able to "identify the meanings of common Elizabethan words" and an inclass essay as a high cognitive way to assess the ability to "distinguish thematic and textual differences" (A3, p. 4). Kochab also chooses points and percentages that appear to align with his goals for this unit.

On the next page, however, when he is asked to verbally defend his point values and weighting, his descriptions are vague. He can explain why some components weigh more than others, but does not really talk about relative weight. He does say, "each outcome is given the
weight equal to the amount of knowledge needed to generate a satisfactory response" which shows that he is at least thinking about weighting (A3, p. 5). It might take more content knowledge, however, to be able to really know how to create weight. Thus, this blueprint and the description indicate that the goal of this assignment is to get PSTs thinking about the processes behind teaching, even if they have yet to really know how it works.

In many places, Kochab continues to highlight that he is still a PST, and has not yet had classroom teaching experience. For example, he plans to assign a take-home essay with one night to complete it, and also on the night before the written in-class exam. This shows that he is not accounting for the real-life influences on his future students, and is not yet aware of how time works for teenagers. Nevertheless, things like student time management come with experience, and Kochab is not yet expected to have learned this. This time management again comes up in his description of how long he expects his exam to take students. He gives them fifty minutes to answer ten matching questions, ten multiple-choice questions, and 10 short answer questions, and to write two essays. However, again, the ADP does not appear to be measuring Kochab's knowledge of student time, and therefore this conclusion, for now, is mostly irrelevant. All it does is highlight that the core of the assessment course is not rooted in the ability of the PSTs to yet be able to accurately understand either the content or the students.

In the second part, Kochab attaches the traditional examination he has created to end his unit, following the instructions from his blueprint. According to the task description, he must include student-friendly directions and format the exam as a student would see it. Over the next few paragraphs, I will take excerpts from this test and come to some conclusions about what we can, and cannot, assess about Kochab from this submission.

The first thing I notice when I look at the examination is the general directions. Kochab writes, "Write a bit more on the back of each sheet if needed" (A3, p. 8). This direction is informal for directions, and missing a comma. Nevertheless, it is clear that Kochab has given some thought into how his students might want to interact with the test and has provided the requisite space to do so. Thus, it suggests that while not yet an expert direction writer, he is taking into consideration how he should format his spacing to allow for student thought, and suggesting that his students take the time and space needed to consider their answers.

Next, there is the matching section, which from the blueprint, will be about understanding Elizabethan English and be measured using low cognitive demand. The first section, however, while low demand, is not necessarily age-appropriate for his target grade. Asking 11<sup>\*</sup>-grade students to match "didst" with "did" seems to be below grade level. Yet, as the ADP is not a contentfocused task, Kochab is not being assessed on how well he knows what content to include for his hypothetical students, and he does have six options for five matching words, which means he has considered how to eliminate guessing from his examination. Here, we can see that, structurally, Kochab is doing as expected. His next matching section exhibits similar results. The word choices might bring in some construct-irrelevance by asking words in the definitions that might not be in their common vocabulary (e.g. using the word "honorific" instead of title), and he might get some calls from parents for using the words "cuckold" and "bastard," but again, he uses six options for five questions, so his knowledge of the structure of matching is sound (A3, p. 9).

In the multiple-choice section, we again see the same pattern. His questions are not evaluated through the lens of a content expert, but regarded as practice writing multiple-choice items. He experiments with different answer lengths, and avoids making the right answer always they longest choice. These strategies fit the guidelines found by Haladyna and colleagues (2002) in

their research about best practices for writing multiple-choice questions. From their research, 85% of the textbooks in their sample recommend making "choice length equal" to avoid unnecessary distractors, and the remaining 15% of the books said nothing (as opposed to being opposed) (p. 314). While, technically, Kochab does not make all his choices the exact same length, by avoiding a pattern, he achieves a similar goal. Kochab puts no more than four items on a page, giving plenty of space for students to think and draw out their ideas. He avoids cluing in his questions, and when the next question is asking about whom Hero marries, the prior question refers to him only as "her betrothed" (which might bring in some bias around vocabulary, but at least shows a concern about clueing) (A2, p. 11). According to Hadalyna and colleages (2002), 95% of textbooks on multiple-choice writing recommend avoiding clueing (and the other 5% said nothing (p. 314). Still, there are some concerns. For example, question 20 has no correct answer, although that is possibly a type-o. The biggest issue is that the exam does not show a deep understanding of the content, but that is not the purpose of the examination.

The other major issue with this examination is the provided rubric for the essays. The three essays each have rubrics that are provided to the students with the essay prompt. Providing these rubrics indicates that Kochab understands the importance of being clear with expectations, and understands that providing weights can help students know what needs to be included. The rubrics, however, are very general. For example, Kochab has "Explanation of how the character's actions help shape the genre of the play is logical and robust" worth 5 points (out of 15) (A3, p. 16). What does it mean to be "robust"? Will students reading this rubric know what is expected of them? Further, when Kochab uses this rubric later to assess his students, does he know what the difference is between receiving 3 points or 4 points? Thus, while providing a rubric, with weights, shows that Kochab has thought about how he would grade this essay, the generic breakdown with

large point values indicates that he has not fully thought through how it would be graded in practice. Of course, the trick here is that it might take reading several student responses until a more completed rubric could be developed. To create a more fine-grained breakdown now would likely lead to a rubric that does not reflect the material presented by the students and might end up attributing points poorly. Thus, from the submission, we can learn that Kochab has a general understanding for how he would grade student essays, but cannot know from the provided work how his rubric would act in practice.

Additionally, the rubric for the Take-Home essay is even more generic than the in-class essays. There are eight points (out of 30) allocated to "The writing is creative and uses new ideas while staying in the realm of realism in the play" (A3, p. 17). It is not clear how these points would be allocated in practice. What Kochab does demonstrate, however, is that despite offering his students two choices for the essay, he has managed to create a rubric that will be usable for both essays. It would need further inspection to ensure that one essay choice is not significantly easier than the other, and that the page restriction for the essay is reasonable, but it still demonstrates that Kochab was able to come up with two options for his students that theoretically could be graded on the same scale.

After the traditional assessment, Kochab adds, as directed, a Supplemental Item Bank (SIB). I had assumed, from the task description, that these items would be additional questions that could be substituted into the traditional examination. Seeing Kochab's work, however, it appears that instead these items are alternative ways to assess the same material as in the traditional test. For example, in the traditional test (A3, p. 12), Kochab asks a short-answer question about how Shakespearean comedies always end, and in the SIB, students are given the statement "The play ends with marriage or the promise of marriage" and are asked to select between comedy and

tragedy (A3, p. 20). Thus, seeing that this was how the SIB was enacted, shows that a different skill is being assessed than I originally understood. The SIB is not for questions that were cut because it would make the test too long, but instead is assessing that PSTs can think of multiple ways to assess the same material. However, with only one sample work, I do not know if this was the same interpretation for all PSTs.

The area where initial thinking and format is prioritized over content knowledge and knowing about students is most easily seen in the final component of Part 2, where the PST is asked to imagine a student with a special need in his class and adapt the examination as appropriate. Kochab goes above and beyond by considering two students with different special needs, instead of just choosing one. As the task description states, this component will require "intelligent speculation" (A1, p. 4). This speculation is clear because to make accommodations accurately, one would need knowledge of special education and proper adaptation techniques. However, the PST at this point in his or progression (or perhaps ever, since it will likely be a special education teacher who makes the actual accommodations for students) may not have the necessary knowledge to know how to make assessment adaptations, since a special education course is not a required pre-requisite to Course C. Therefore, the PST needs to just make their best guess, and likely be judged on effort, not outcome. Ways that I can see that Kochab is making logical guesses is in how he makes his adaptations. For example, he gives his student with ADHD time and a half. When explaining his thinking, Kochab explains that this is enough time for his student to finish, but not so much time that it would be an unfair advantage. The explanation itself shows care and concern, but the calculation appears to essentially be random, even if commonly used in schools. Nevertheless, Kochab exercises thought, and is rewarded for it.

Another example of privileging thought and effort over accuracy is when Kochab says he will have his student take the exam in a room without other students and without a window (A3, p. 21). This might be helpful for the student, or it might not. Kochab demonstrates that he does not yet know about accommodating examinations, but he is offering a suggestion. For his other hypothetical student, he offers to give her the text one page at a time to minimize stress. While this seems like it will be a good idea, one then quickly realizes that he is now having her take the test in a quiet room, one page at a time, and then must come back to the main room to get each next page. Would this really be less stressful? I do not know, but I am not confident it would work. I would need to consult her special education teacher and read her IEP to be sure.

Finally, Kochab once again reveals that he is unaware of the potential time constraints of a school. He says he will proctor his student with ADHD during his time and a half in a separate room. While it would be great if all teachers could proctor extra exam time to offer support and answer questions and keep students on task, the logistical constraints of doing so while also having a full course load of other students may not be possible. Nevertheless, Kochab is showing that he is thinking about his potential students and demonstrating a willingness to think about accommodation, and for that, he receives the A.

In Part 3, Kochab designs an alternative assessment for his students. He decides to have the students write and perform a skit based on the characters from both plays. It is an interesting idea, as it has the students consider both of the plays and their characters, and then find a way to mix the two genres. It is likely, as Kochab claims, that the task uses a high level of cognitive demand. Interestingly, in an attempt to focus on the content and not the performance aspect, Kochab reiterates several times that the performance itself (other than length of time and equal participation) will not be graded. The purpose of this component appears to be to get the PST thinking, and, as such, it is not fully graded on what comes from that thinking. I also guess that part of what sets Kochab's submission apart from the other PSTs, is that he, unbidden, gives adaptations to this alternative assessment for his two chosen students with special needs. This adaptation is not a required element, yet Kochab does it anyway. This demonstrates that once in the mindset of the classroom teacher, Kochab is thinking about all his pretend students.

The big takeaway from Part 4, the reflections, is that Kochab knows how to "talk the talk." He is able to express that he learned and grew and can point to specific components of what he learned. For example, he states that he "understand[s] that no one assessment tool will be able to cover a whole classroom without some type of adjustments being made" (A3, p. 30). Whether this is something that was actually learned through the experience or it was something mentioned in class that sounded pertinent, Kochab's inclusion of this statement at least highlights that he knows that this point is important. Assessments are a great tool for teaching, but there is no magic solution. Kochab also mentions other important teaching skills, such as revision and reaching out to colleagues for help.

From his second reflection, where he talks about the College of Education goals, it becomes apparent to me that class time was used to develop the ADP submission. Kochab references justifications to his choices that do not appear to be present in the submission. I am left to assume that as he developed his project, he worked with his peers in class and made adjustments and defended his work where necessary. He talks about "realistically" mapping out his test. From what I saw in his length description, it did not appear to be realistic, but if he claims it is, then I am guessing he spent class time trying it out. Penultimate, he mentions that this project made him pay attention to formatting, which leads me to believe that Dr. Polaris did pay close attention to the format of the designed test, even if he did not read the content questions. Lastly, Kochab adds the final line describing what he learned from this experience: "instruction needs to be aligned with assessment" (A3, p. 33). How exactly writing a test and adapting it leads to this conclusion is not entirely clear, but he claims to have learned it.

# Interview with Dr. Polaris

In this section, I take excerpts from my interview with Dr. Polaris to better understand what, according to him, is and should be assessed in this course. All quotes are from either my notes or the transcript and will be noted accordingly.

When asked to describe the course that he teaches, Dr. Polaris replied that it was a course in assessment and evaluation that looks "mostly from an academic point of view." By this he clarified that the course was not looking at assessment designed specifically for "special education, or motor skills, or affective tests," but that the purpose was to get PSTs thinking about how to evaluate academic knowledge in the classroom (interview notes). He says that most of the course is "technical." By this he meant that there is a focus on writing traditional and alternative assessments and how to use them in the classroom. Thus, the assessments in this course should align to this purpose, and it is not surprising that the core assignment, the ADP, is assessing the PSTs' ability to write both.

Dr. Polaris expects that the PSTs who enter his classroom already understand what it means to be a teacher and have a philosophy of education. Also, since the curriculum course (Course C) is either a pre-requisite or a co-requisite course, he expects that by at least mid-semester his students know how to write objectives, use standards, and create student outcomes. He also expects that the PSTs in his classroom know how to be students and can demonstrate scholarly behavior. They do not need to know anything about assessment, but they should know how to learn. Thus, his course covers assessment knowledge and writing, with only a minimal portion of the course focusing on how to be a teacher.

I asked Dr. Polaris what he expects that his students will know and be able to do by the end of the semester. Dr. Polaris picked up his syllabus and read me the five essential outcomes, and then handed me a copy of the syllabus. It is interesting to note that he was the only professor whom I interviewed who made this link as explicit. Thus, what he says he wants students to learn and what he tells the students they will learn is wholly aligned. Essentially, he believes that in the end, his students will "know assessment really well." He backed this claim by stating that their recent CAEP accreditation had mentioned how well the students at this College of Education knew assessment. Interestingly, in addition to listing the outcomes, Dr. Polaris also added a bit of commentary. What stood out to me, was that he added, when talking about traditional tests, that there was "nothing wrong if done right." This makes me think that in addition to teaching his PSTs how to write and analyze assessments, he will also share this view with his students that tests are not bad, despite what the PSTs might hear on social media, in the news, or from other outside sources.

Dr. Polaris and I talked about the assessments he gives in this course and how PSTs get their grades. During our discussion, Dr. Polaris told me that the ADP was worth 40% of the grade. The syllabus says 50%. This discrepancy may be due to spoken error, or perhaps, because he has taught this course so many times, 40% might be what the assignment used to be worth. Regardless, the ADP is worth a considerable amount of the grade. Dr. Polaris also added that if students can write a summative assignment, then he assumes that they can write minor assessments. This claim influences how and what he grades. In terms of weighting, he attempts, to the best of his ability, to align the weights in his course to what he finds most important. Speaking of grading, Dr. Polaris said that he does not give full rubrics, or as he says, he is a "a rubric minimalist". He believes that when rubrics are detailed, they may not only freak out students, but they can limit both creative thinking and the teacher's professional judgment. He believes that rubrics are intended to show the students what the teacher is looking for, but should not be overly prescriptive. He says that if you want to include details, only put the top requirements and then say as those are not met, you will lose points, but he does not think that "richness" can be captured in a rubric. This explains both why the rubrics on his task descriptions are so vague, and why Kochab wrote general rubrics as well.

Dr. Polaris also says that he grades in a way that he hopes will reflect the mastery of the goals in the course. PSTs are given "ample opportunity" to review submissions with him before submitting them. He encourages the PSTs to meet with him early on and he will give feedback. Major assignments do, however, have "drop dead dates" where after which no submission will be accepted.

Both the ADP and the AP are departmental expectations, and thus Dr. Polaris assigns them. The two tasks are not some surprise assignment that has been handed down from on high. Instead, they are "agreed upon" by the department and its members. The weighting of these assignments, however, is up to him. He has found, furthermore, that if the students can do well on these two assignments, it is likely that they can also do well on the other components of the course, like quizzes and tests. So, even if not much room is left in the rubric for the other components, it is expected that the students can do them. And, while there are no official rules for the weighting of the two departmental assignments, Dr. Polaris said that if you weighted it around 5%, "your colleagues would be all over you." Thus, there is academic freedom, but within reasonable limits.

#### Looking across

As informative as it is to look at the four indicators separately, much can be learned from considering them together. Looking across, I was able to build a matrix that matched the course assignments and the essential course outcomes as stated in the syllabus. To build this matrix, I started by putting the objectives as the columns and the five graded elements (midterm, final, AP, ADP, and professionalism) as the rows. Then, I broke down the AP and the ADP using the rubrics. Because there was no weight assigned to the components of the AP and the ADP, I gave each component equal weight. Then, using the information from all the four indicators, I filled in my chart. Also, because I had no real information about the midterm and final other than what was stated in the syllabus, I decided that it was likely that each would balance the three objectives most easily assessed in paper-and-pencil format. Because I had no real information to help me support this further, I divided the weight equally.

As I built my chart, I realized that I needed to add two columns in addition to the five objectives. One column was for General Mechanics and rule following. Often in the rubrics there was a space for following the format or including all the proper components. Also in here one sometimes found references to spelling or grammatical errors. Because this was so prevalent, I felt it was necessary to include grading weight specifically for this. I also found that there were elements assessed in the projects that did not directly relate to any of the *5* essential outcomes. However, because these were varied, I labeled this column as "other." The chart is posted below (however, all objective names and rubric elements have been replaced with letters and numbers for IRB security. As with Course C, I had Katie Cook help me conduct reliability tests.

Table 6.1 Matching Course Components with Course Objectives												
Assignment	Total Weight	Component	1	2	3	4	5	General mechanics & rule following	Other			
Midterm	10%											
Final Exam	15%											
AP	15%	a.							1			
		b.							1			
		с.		1		1						
		d.						1				
		е.				1						
		f.							1			
		g.						1				
ADP	50%	1a.						1				
		1b.							1			
		1c.	1									
		1d.						1				
		1e.			1							
		1f.						1				
		2a.						1				
		2b.			1							
		2c.						1				
		2d.			1							
		2e.			1							
		2f.	1									
		2g.						1				
		3a.						1				
		3b.						1				
		Зс.			0.5							
		3d.			1							
		3e.						1				
Professionalism	10%	а.					1					
		b.					1					
		с.					1					
		d.					1					
Totals	100%		2	1	5.5	2	4	11	4			

In this format, the chart, while interesting, did not lead to any apparent conclusions because, using the number 1, everything was weighted equally. I soon realized, however, that if I transposed my matrix and added weighting, I could recreate the course as it was designed through assessment.

Table 6.2 Transposed Matching of Course Components and Course Objectives																																
Assignment	With weight	Midterm	Final Exam	L	AP ADP														Professionalism													
Total Weight		10%	15%				15%												50	)%									10%			
		10	15	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.78	2.50	2.50	2.50	2.50
Component				a.	b.	с.	d.	e.	f.	g.	1a.	1b.	1c.	1d.	1e.	1f.	2a.	2b.	2c.	2d.	2e.	2f.	2g.	3a.	3b.	3c.	3d.	3e.	a.	b.	c.	d.
1	13.9%	0.33	0.33										1									1								1		
2	10.5%	0.33	0.33			1																							<u> </u>			
3	15.3%														1			1		1	1					0.5	1			1		
4	12.6%	0.33	0.33			1		1																								
5	10.0%																												1	1	1	1
General mechanics & rule following	\$ 29.3%						1			1	1			1		1	1		1				1	1	1			1				
Other	9.2%				1	1			1	1		1																				

Zooming in to what is important, one can see how the grade is calculated based on which elements

are assessed where:

<b>Table 6.3</b> Weighting ofEssential Course Elements											
Essential Element	Weight in Course										
1	13.9%										
2	10.5%										
3	15.3%										
4	12.6%										
5	10.0%										
General mechanics & rule following	29.3%										
Other	9.2%										

Nearly 30% of the course grade comes from General Mechanics and Rule Following. Thus, this course, which does not put this as an essential objective, nevertheless gets the bulk of its grade from this element. This weight comes from the sheer number of times this comes up on the rubrics of the **AP** and the **ADP**. Each portion of both assignments has this in some form on its rubric. It is possible that the actual weight each time is lower, but this still shows the point that without following the rules or using general mechanics for the English language, one's grade in this course would suffer.

With the second highest weighting, after mechanics and rule following, is outcome 3. This outcome, that students can construct quality assessments, is unsurprisingly highly assessed. From my conversations with Dr. Polaris, this stood out as critical for the purpose of this course.

Furthermore, since the ADP is designed around this element, it is apparent that the department as a whole sees this outcome as essential. It is interesting that this key outcome is listed third and not first, if it is the most important. However, as hypothesized earlier, it might be that outcomes 1 and 2 are precursors to outcome 3, and thus the first three outcomes are listed in the order they will be covered in the class.

Interestingly, with equal amounts of least weight, at 10% each, are professionalism and being able to meaningfully critique assessments. Professionalism being 10% is not that surprising, since it matches what was said in the syllabus. Critiquing also being only 10%, however, is surprising because the **AP**, which is dedicated this essential outcome, is worth 15% of the grade. However, this means that one third of this project does not actually assess the core of the assessment. In fact, when considering just the rubric for the **AP**, only one of the 6 rubric elements attends to critiquing. The rest of the weight attached to this element actually comes from the midterm and final exam.

Another major thing that stood out from looking across all the indicators was that this course is not discipline or subject specific. This matters because it means that the assignments and projects are not graded based on plausibility of being actually usable in a classroom. Because the PSTs are still early in their program and have not yet taken any subject specific methods courses, the purpose of this course is to understand the structure of assessment writing and analysis. This leads to an interesting outcome, because it means that PSTs can receive full marks for designing an assessment or supposing why a student got a question wrong without being "right." The purpose of this course is not to ensure that PSTs can make actual usable tests, but that they know the structure so that when they learn more in their methods course, they can apply what they learned from here into it. Similarly, the PSTs do not need to actually know why the students got questions wrong, but

they need to know how to look for errors and have experience coming up with hypotheses. It appears that the theory is that this knowledge will be transferable, and will improve with methods course experience. In future projects, it would be interesting to see how this plays out in reality. One way to do this would be to analyze tests written by these **PSTs** once they have completed their methods course, and to interview them about their philosophy and approaches toward assessment once they have taken both courses.

### Summary

Overall, Course A prepares PSTs to be able to design and critique academic course assessments. The assignments, both core and otherwise, prepare PSTs for this goal, providing practice and guidance. Additionally, Dr. Polaris emphasizes professionalism and rule following, requiring that the PSTs not only learn the course material, but also to act and present their work in specific ways. By the end of the course, PSTs should have a strong introduction to assessment and be able to use what they learned in future courses.

### Chapter 7: Course BD

"Do I scale back or do you bump up?" - Dr. Altair (transcript)

"I think we have a good sense of what our students can do in our course, but overall as a program, it is hard to say because we are all different." - Dr. Deneb (transcript)

### Overview

Course BD is a technology class for PSTs in the middle of their program, although it is a bit different from the other two courses I studied at Galaxy University. While it still takes place in Phase 2, which is after general education courses and before method courses and student teaching, this course is currently taught by multiple instructors and in different forms. In the semester when I conducted my research, Dr. Altair was teaching a fully online version of the course, and Dr. Deneb taught a hybrid version. There was a third instructor, but he was not part of my study.

For continuity, all documents related to this course are marked by the letter **B** or **D** and a number. All documents associated with **D**r. Altair get the **B**, and have names from the Aquila constellation. All documents associated with **D**r. Deneb are denoted with a **D**, and have names from the Cygnus constellation.

Because there are two separate instructors, there are separate syllabi, assignment sheets, and rubrics. However, because they are still connected to the same course, I analyze the documents all in this same chapter. As I work through the sections, I highlight the key points from each document and include a summary section comparing and contrasting when appropriate.

It is important to note that while the interviews were conducted in one semester, the professors for both versions of the course gave me their materials from a different semester. Thus, unlike the other courses for which information is all contained in one semester, Course BD contains information from three. Nevertheless, since the course is still taught by the same instructors, the variation in the data should be minimal.

### Syllabi

#### Dr. Altair's syllabus.

Possibly because this is an online course, the syllabus for Dr. Altair's section of Course BD is quite extensive. Her syllabus is posted on her course website, and is several website pages long, which translates to even longer PDF pages. As I describe her syllabus, I will cite each page both by website page (B1-B8) and by PDF page length within each website page (page number). Combined, the name will follow the format of "BX, p. x." There is some repetition in the syllabus that is likely due to the website nature. Unlike a paper copy that is normally read from front to back, websites are more fluid in their order and thus it may not be expected that PSTs will read every page in detail.

Before the syllabus begins, Dr. Altair starts her website with an introductory page welcoming her PSTs to the course. After her general welcome, Dr. Altair gives a short written introduction to the course. She explains that the course covers "teaching" and "learning" and both words are bolded for emphasis (B8, p. 1). She also bolds the full sentence saying that the syllabus may be "changed or altered" before the course begins (ibid). She then continues with a paragraph about general expectations for an online course and recommends that PSTs who have never taken an online course before follow a link to the university's description of how the online system works.

Next, Dr. Altair copies the description of the course as written in the course catalogue. Presumably this is the same description regardless of which section the PST takes. She then lists the courses that are pre-requisites for the course, which include Course C and admittance to the

teacher education program, and then gives a short biography about herself. Her biography is professional, mentioning her research interests and experiences. She concludes this page with a photograph of her face, which potentially is to give a human aspect to this online course.

The first page of the syllabus begins with a more thorough prose description of the course. The adverbs used in the introductory paragraph are "critically and creatively," suggesting that PSTs are expected to take an active role in their education for this course (B1, p. 1). In the next two paragraphs, Dr. Altair emphasizes key words and ideas with bolded text. She emphasizes "practical examples" and "discussions of the pedagogical and ethical issues" (ibid). Thus, this course is not just about theory, but also has a strong focus on practice. Dr. Altair also emphasizes that this course is "not" about merely learning how to use technology or about becoming prepared to teach technology use to future students (ibid). Instead, the purpose of this course is to learn how to integrate technology into the teaching practice.

The second page of the syllabus details the course standards. Dr. Altair explains that this course is based on the International Society for Technology in Education's (ISTE) National Educational Standards for Educators, as well as the state's professional standards for teachers. She then lists the standards word for word, and concludes with a statement that PSTs will demonstrate that they have met all these standards by submitting a web-based Summary Portfolio (SP). Dr. Altair adds an extra comment at the bottom of the page warning PSTs that this course is computer-based, and that if they are not familiar with the computer or its programs, they "will need to allow extra time" (B2, p. 3). In fact, her entire warning is bolded for emphasis. She also reminds PSTs that if they have not taken the pre-requisites, they will not be permitted to take the course.

The next web screen is allocated to detailing the resources that will be needed for the course. As stated in bold at the top of the screen, there are "no required textbooks" for the course

(B3, p. 1). Instead, PSTs should be prepared to use the course website and the Internet to find and view all required readings. Dr. Altair offers an optional textbook if the PSTs want something more concrete, but buying this book is not required. She also suggests that if the PSTs are more comfortable reading printed materials, they should acquire a binder to keep all printed versions of the electronic text. She warns, however, that PSTs should not print out material too far in advance because material is likely to change.

After the text information, Dr. Altair lists 16 bullets that contain information about required electronic devices and software necessary for the course. Websites and plug-ins are listed as hyperlinks with short descriptions of how or why they will be used. For example, one bullet says: "Canvas Conference, Skype (http://skype.com/) or Google Hangout

(http://www.google.com/+/learnmore/hangouts/) for online Office Hours - your choice" (B3, p. 3). Thus, PSTs are informed that office hours will be held electronically, and that they have three options for how to attend. Another bullet informs PSTs that they will need to create a twitter handle and follow Dr. Altair. Interestingly, PSTs are told that they will need also to have course notifications sent to a smart phone. It is not clear if the notifications will contain the same or different information from the tweets, and if having a smart phone will be a real requirement for the course. Other than the smartphone, all the other bullets contain links to and descriptions of materials that should be free, or at least relatively cheap, to acquire.

The subsequent web screen is dedicated to "Policies and Expectations" (B4, p. 1). The first line states that all work "must" (bolded for emphasis) be completed using a computer (ibid). PSTs are informed that if they cannot get an application to run on their personal computer, they may use the computer lab in the Education Building on campus or in the university library, and hyperlinks to both locations are included. She adds a bolded note that the people who work in the computer lab are there to help with "technical difficulties only" (ibid). They are not there to help with the content of the course material nor are they there to teach the PST how to use the software. For emphasis, she bolds that PSTs should, "please not ask" the lab workers for this kind of help (ibid).

The next policies are about participation. In terms of online discussions, Dr. Altair explains that the course discussions are to be "student-lead" *[sic]* (bolded for emphasis), although she will facilitate and clear up any misconceptions (B4, p. 1). Dr. Altair also explains that there will be group projects that all PSTs will be required to complete. She acknowledges that group work in an online setting may be challenging, but that to "participate fully and contribute" (also bolded) is both expected and required (ibid). PSTs are expected to be professional in their collaboration, and the syllabus states that learning to be collaborative is "critical" for future teachers (ibid). At this point in the syllabus, how these discussions and projects are to be graded is not mentioned. She does mention, however, that there will be no exams in the course. She also says that she will return all work with feedback two weeks after it is due.

Web screen five of the syllabus is dedicated to general course requirements. Throughout this screen, Dr. Altair bolds key words to help focus attention and highlight key points. First, Dr. Altair calls attention to the course website and encourages her PSTs to pay regular attention to the announcements posted there. She warns the PSTs that as the course progresses, there will be more information to wade through so they should be active in their scrolling and can use a service to update them of new material. Second, Dr. Altair reminds PSTs once again that the syllabus may be revised throughout the course, and that any changes will be announced. If the change is to a due date, she will be sure to make the announcement with enough time to react. Third, Dr. Altair mentions that as an online course, the PSTs will have considerable responsibility for staying on

track. She warns the PSTs that there will be a substantial workload for this course, and that they will not want to find themselves falling behind.

After the general notes, Dr. Altair dedicates space to describing what it means to be a successful online student. She provides two bulleted lists describing what "experts" believe to be traits of successful online students (B5, p. 1). In addition to the lists, she includes a citation to the experts, and a link to a readiness quiz that the PSTs can take on their own time to see if they are well positioned to take this version of the course.

There is next a list of ten bullets that describe the "learner requirements" (B5, p. 2). Many of these bullets include items that have already been mentioned in the syllabus (like backing up files, working in teams, and checking the course website), but here these requirements are bunched together and explained in some more detail. For example, instead of just telling students that they need to check the course website, she says that it should be checked "daily" and daily is bolded (ibid). There is some new information provided as well, such as information for **PST**s with documented disabilities on how to get course accommodations, and notes for international students.

Finally, the page concludes with a green box about academic dishonesty. Dr. Altair states that all forms of academic dishonesty "including all forms of cheating, falsification, and/or <u>plagiarism</u>" will "not be tolerated," with the second quote being bolded (B5, p. 3).

Web screen six details the assignments that will be part of Course BD. Dr. Altair begins with a paragraph about timing and due dates. First, she reminds the PSTs again they should reference the course website to find all their readings and that due dates can be found there. She explains that despite this being an online course, which will offer a bit of flexibility, PSTs should still plan on meeting assignment deadlines. She tells her PSTs to not plan on turning in the bulk of

the assignments at the end of the course, and suggests they should instead budget their time throughout the semester. She also explains that modules will be "released in groups" throughout the semester (B6, p. 1). PSTs should not expect to be able to see everything about the course from Day 1.

After these introductory warnings, Dr. Altair has a section on assignments. She begins by informing PSTs that all assignments will be due at midnight on their due date. She also reminds students that all assignments must be submitted into the correct dropboxes, and that she will not accept emailed submissions. Additionally, she says that some assignments will be short and technical, while there will also be "longer, more complex activities" that will allow the PSTs the ability to show off their new skills (B6, p. 1).

Next, Dr. Altair lists all the assignments that will be included in the course with short descriptions. The assignments are:

- Designing "technology-infused" lessons. These lessons can be built off lessons designed for other classes as long as they are adapted to meet the new requirements and the PST has asked permission from Dr. Altair first (B6, p. 1).
- 2. Creating several lesson components that use technology. These components are things like videos, handouts, and web pages.
- 3. Contributing to the online discussion forums. **PSTs** will be graded on both quantity and quality of the contributions.
- 4. Researching technological tools that will likely be used in their future classrooms.
- 5. Participating in a virtual showcase.
- "Developing content to support student learning," which will be done as a collaborative group project (B6, p. 2).

- Creating an online lesson that will be used to teach peers, as well as completing online lessons designed by peers.
- 8. Writing reflections in response to given prompts.
- 9. Creating an online, digital Summary Portfolio that encapsulates all that was learned during the semester. This SP is the core assignment in this course, and as such, requires a C grade in order to pass Course BD.

While the assignments are listed and enumerated, no point values or weights are given at this point.

Grading is given a separate heading on this web screen. While there are no exact details still about point values or weights, Dr. Altair does explain that these details will be provided on the course website within the different modules. She explains, however, that some assignments will be credit/no credit, while others will be graded more thoroughly. She says that, "to receive credit for an assignment, the artifact must be **technically** correct **AND** have appropriate **pedagogical content** (B6, p. 2, bolded as presented in the syllabus). Additionally, Dr. Altair comments that PSTs should expect to turn their best work in from the start, as revisions are seldom accepted. For credit/no credit assignments, Dr. Altair informs the PSTs that despite the lack of numerical grade, it is still expected that the assignments be completed. Lack of submission can cost a PST up to 2 points (although what these 2 points mean in practice is not explained). Also, assignments get a "10% point deduction" for each day it is submitted late, up to three days late. After three days, the assignment may not be accepted, although it is not stated that it will definitely not be accepted, so there is some ambiguity left in this statement. Finally, Dr. Altair says that "**the third late assignment will receive 0 points**" (B6, p. 3). Thus, it is clear that Dr. Altair expects that

PSTs turn in their work on time and failure to stay on track will lead to a significant decrease in grade.

At the bottom of the screen, there is a green conversion chart from numerical to letter grades. While similar to Course A and Course C, the chart is not identical. What is notable about this chart is that it stops at 70%. Anything below a 70 counts as an F for this section of Course BD. This makes sense since one needs 70% in this course to get College of Education credit for it. There is no explanation of how rounding will work in this course. Academic honesty is again mentioned and PSTs are warned that not submitting original work will be dealt with accordingly.

The last page in the syllabus is mostly about communication. Dr. Altair provides methods for reaching her directly, through email, phone numbers, office hours (both in person and digital), and twitter. She reminds the PSTs, however, that they are expected to communicate professionally. In terms of professionalism, Dr. Altair writes two paragraphs on how PSTs should communicate professionally with each other, as the course will include many PST-PST conversations. She reminds them that written words are often harder to convey meaning through, and thus they will need to be careful in what they say. She also says that professionalism means working evenly in their groups and "carrying your own weight" (B7, p. 2). Most importantly, however, she bolds that, "All are entitled to respect" (ibid). Thus, professionalism, even if not one of the course standards, still holds an important role in this course.

#### Dr. Deneb's syllabus.

Dr. Deneb starts her syllabus with a short description of Course BD. She also includes a bolded and red-text sentence emphasizing that PSTs must score a C in this course in order to get credit for the College of Education. She states that failure to attain this grade will lead to the PST needing to repeat the course. After the introduction, Dr. Deneb lists the course objectives. These objectives are the same ITSE Standards as used in Dr. Altair's section, excluding the added state standard. She provides a hyperlink to the standards so that PSTs can read the standards in more details.

In the next section, Dr. Deneb provides a conversion chart from numerical to alphabetical grades. Unlike Dr. Altair, Dr. Deneb gives percentages for the D range, as well as all the other letters. Dr. Deneb also tells PSTs to look at the course calendar (elsewhere in the document) to see the list of assignments and their due dates. If they want more details, they can consult the course website. She adds a note that work cannot be resubmitted for a new grade unless they have previously asked her for permission.

In the subsequent section, Dr. Deneb covers her course policies. The first policy she discusses is about academic dishonesty. She writes in bold that Galaxy University has a "zero tolerance policy" for this behavior and hyperlinks to a site (D1, p. 2). The hyperlink is no longer active, but it is likely because the syllabus I am analyzing is not for a current term. She then states that she will be using "plagiarism detecting software" and gives a list of the various ways a person may commit academic fraud (ibid). As consequence for academic dishonesty, she says that PSTs will receive a failing grade in the course, in accordance with university policy.

Her next policy is written both in bold and in green text for emphasis. This policy is that all work must be submitted by the start of class on the day it is due. Unbolded, she adds that it is the PSTs' responsibility to use the course website to stay up-to-date about deadlines. Both sections reference the same platform for accessing the course website, but it seems to be individualized per section. Dr. Deneb then rebolds and uses green text again to state that any work submitted more than twelve hours late may not be graded, or if it is graded, points will be deducted. This rule, while emphasized, appears to give Dr. Deneb a bit of flexibility with how it will be enforced.

The third policy is also bolded and written in green text. It is that PSTs should "schedule about 6-9 hours per week for this course on the computer" (D1, p. 2). In plain text, she adds that handwritten work will not be accepted.

While not given its own heading, the syllabus then transitions to talking about course assignments. Dr. Deneb explains that PSTs should plan on completing one or two modules most weeks. She states that modules will be released a bit in advance of the week so that PSTs can get a head start on the work, but not so far in advance that it interferes with their ability to learn at a reasonable pace. Additionally, she adds that some assignments will be due at 11pm on Tuesdays, while others will not. All due dates and times are listed in the calendar section of the syllabus. Dr. Deneb also includes a short note about keeping materials. She tells her PSTs that they need to have access to the materials they wrote all semester long, and therefore recommends using a flash drive or Google Drive.

The next section is dedicated to course materials. PSTs are to have access to a computer that can access the Internet. They need to activate a special account that will allow them to submit their SP. A hyperlink to the activation site is provided. PSTs also need to have access to Microsoft Suite or its equivalents, including a document builder, a presentation builder, and a spreadsheet application. A hyperlink is provided to reach OpenOffice. Lastly, PSTs need to have Adobe Acrobat Reader and Dr. Deneb provides a link to download the free desktop version, and she mentions that there are also phone and tablet applications that they can download if they want.

The syllabus concludes with a detailed calendar for the semester. Dr. Deneb lists class times, when to start modules, and when assignments are due (both date and time). From this calendar one can see that the assignments will be 14 modules, a tech survey, several online discussions, a telecollaboration, a WebQuest, a proposal, a lesson plan, a Moodle, a showcase, and a portfolio. Unlike Dr. Altair's syllabus, however, summaries of the assignments are not provided here.

#### Looking across the syllabi.

Dr. Altair has a considerably longer syllabus than Dr. Deneb. It is likely that this difference is due to the fact that Dr. Altair's course is taught entirely online, while Dr. Deneb teaches a hybrid course. With the additional face time, Dr. Deneb likely can describe her course verbally to her PSTs. Also, with this course being the first online course for many PSTs, Dr. Altair has to add additional information about how to take and participate in an online course that Dr. Deneb is able to skip.

Drs. Altair and Deneb give very similar descriptions of their courses with only two differences. Dr. Altair says the course will "help" PSTs to "critically and creatively apply the concepts, principles, hardware and software associated with the infusion of technology in solving educational problems and meeting challenges" while Dr. Deneb instead says the course will "guide" the PSTs (B1, p. 1 & D1, p. 1). Also, Dr. Altair says that the PSTs will have the "roles as a classroom teacher" while Dr. Deneb calls these the "roles as facilitators of learning" (ibid). This suggests that they are describing the same course, but attempting to use their own words. It is interesting that Dr. Altair uses more traditional wording ("help" and "classroom teacher" while Dr. Deneb uses more inquiry-based vocabulary.

Both instructors put a heavy emphasis on academic honesty. Both provide links to the university policy and say that PSTs who plagiarize may fail the course. Only Dr. Deneb, however, says she will be using detection software.

In terms of deadlines, both professors tell their PSTs when assignments will be due and put the responsibility on the PST for getting work in on time. Dr. Altair will allow her PSTs to turn in

two assignments up to three days late, while Dr. Deneb will only accept work up to twelve hours late, but does not give a cap on how many times this can happen. Both inform PSTs that they should plan on submitting their best work the first time, as revisions are likely not accepted.

Unsurprisingly, Dr. Altair's class has a much larger list of required digital materials. As a fully online class, there needs to be software and hardware that will allow for this course to replace the face-to-face meetings of Dr. Deneb's course. Both professors, however, appear to have put thought into how to keep costs low for students. Only Dr. Altair provides written instructions for how to access computers, but again, her syllabus needs to be longer since she does not have the inclass time to explain her policies. They both, however, strongly recommend acquisition of a flash drive (or equivalent) to keep copies of all submitted work.

A final difference is that Dr. Deneb lists the full semester's worth of assignments and their due dates in the syllabus and Dr. Altair does not. Instead, Dr. Altair sends PSTs to the course website to see this information. This difference may relate to Dr. Altair's comments about how the course may change throughout, while Dr. Deneb does not provide this caveat.

### Major course assignment descriptions and their rubrics

For both sections of Course BD, the core assignment is the Summary Portfolio (SP). This assignment is departmentally designed and given to all PSTs taking Course BD. The exact directions, however, are adjusted at the discretion of the individual course instructors. To pass the course, PSTs must score a C on the SP, but what counts as a C also may vary by instructor.

In general, the SP is a web-based project where the PST must summarize what they have learned about technology and learning with respect to the course standards. From looking at both task descriptions, it appears that the PSTs design a website with webpages dedicated to each of the standards. On each page, the PSTs must describe and explain how they have met the standard, and provide examples (often with live hyperlinks) of how they have demonstrated their learning. The website is to be professional and well-made, and grading depends on both the quality of the content as well as the presentation.

Dr. Deneb and Dr. Altair give their PSTs identical rubrics for the project, even though their directions differ. There are 23 possible points for this project. Twenty-one of the points are divided evenly among the seven technology standards. For each standard, a PST can score Unacceptable (1 point), Acceptable (2 points), or Target (3 points). In general, Unacceptable is described as having a basic awareness or understanding of the standard, Acceptable is demonstrating some knowledge of the standard and how to use it, and Target is using the standard in a meaningful way. Each standard, however, is given its own description with levels pertaining more specifically to the point of the standard. In Figure 7.1, one can see (from left to right) descriptions of the standard, unacceptable, acceptable, and target. I have highlighted the key differences using color and underlining.

II. Leader: Educators seek out opportunities for leadership to support student empowerment and success and to improve teaching and learning. Detailed Performance Indicators Exhibits awareness of digital tools and their uses in a global and digital society.

Exhibits the knowledge and skills to <u>participate</u> as an engaged professional in a global and digital society. Exhibits the knowledge, skills, and attitudes of an innovative professional in ways that <u>embrace</u> empowered learning with technology, advocacy for equitable access, and serves as a model for colleagues

## Figure 7.1 SP Rubric Excerpt

The remaining two points in the rubric are for professionalism, defined as having correct spelling and grammar, as well as using a layout and design that enhances the project. Despite the identical rubrics, however, the cut-point for passing differs. Dr. Altair states that PSTs must earn 14.5 points to pass the SP (B10, p. 1). Dr. Deneb, on the other hand, requires that PSTs earn 16 points to pass.

It appears that both instructors provide a template for their PSTs to follow while

completing the project, although only Dr. Deneb gave me a copy of hers. Both instructors gave me

sample works, and from a quick overview of the samples, it appears that their templates were likely quite similar, although it seemed that Dr. Deneb separated out the description of the standards from the examples, while Dr. Altair asked her **PSTs** to weave the two together.

One difference in the task descriptions is that Dr. Deneb offers her PSTs the chance to write a sample reflection narrative before starting the project. She tells her PSTs in the task description that if they write a sample, they can give it to her for feedback. Another difference is that it appears that Dr. Deneb releases the task description much earlier in the semester than Dr. Altair does. These opportunities for starting early and getting feedback might contribute to why Dr. Deneb has a higher cut-point for passing.

### Dr. Deneb's template.

At the top of the document, Dr. Deneb tells her PSTs to save this document as a copy in their drive. Next, she gives instructions in blue font, and tells her PSTs that wherever they see blue in the text, it will be for directions only. It appears that what the PSTs write in this template will be the text to copy into their website when they get to that point in the project.

The portfolio is to start with a cover page that talks about who the PST is. They are to put their name, their major, and their desired grade level. Then, they are supposed to talk a bit about themselves, focusing on their teaching philosophy. Also, they are to give a brief overview of their portfolio and what it contains. Dr. Deneb explains that the PSTs do not need to fill a page, and should only write what needs to be included. She also tells the PSTs to include a professional photo in their cover page. This can either be a headshot or a cited graphic.

Next, the PSTs are to dedicate a page to each of the technology standards. For ease, Dr. Deneb includes the standards with all the sub-standards in the template. She tells the PSTs that they can copy the standards directly; although they will need to bold which sub-standards they will

talk about with their artifacts. Additionally, although it is only written once, **PSTs** are to talk about the standards in their own words, as well as copy the exact wording.

Then, for each standard, the PSTs need to have two sections. One section is about what they learned about the standard and how they have grown. Dr. Deneb suggests labeling this section "What I did to accomplish this" (D3, p. 3). In the second section, the PSTs are to include hyperlinks to examples of work they have done this semester that highlights these new skills. Alternatively, Dr. Deneb tells her PSTs that they can integrate these two sections, weaving the examples into the prose.

After listing all the standards and providing place to develop the standards, reflections and narratives, Dr. Deneb provides her PSTs with a table of nearly all the assignments covered in class during the semester. She does mention that this is the same template for all her sections, so some assignments may be extra or missing, and they should modify the list accordingly. In the table, next to each assignment name, Dr. Deneb provides columns to help the PSTs organize their work. One column is for which standard they want to connect to, one is for whether they have included it in their narrative already, and one is for the document name as saved on the PST's personal device. The stated purpose of this table is so that PSTs can stay organized throughout the semester and more easily complete their SP at the end of the semester. Dr. Deneb does add that many of the assignments in the course will meet more than one standard, and thus it is up to them to decide how to best make the matches, and this will depend on how they wish to talk about what they learned from any given task.

Finally, Dr. Deneb gives a brief overview of how the PSTs should write their final reflection. She suggests including three headings, "What I've learned," "What I still need to accomplish & ideas for doing so," and "What is important if I am going to integrate technology

successfully" (D3, p. 10). She then states that this reflection is a "key element" of the SP and that it should be kept brief. Thus, while PSTs should think hard and carefully about how to answer these three prompts, brevity is still important.

As a last note, Dr. Deneb warns her PSTs about copying from the template into the website. She recommends de-formatting all the text first, because otherwise the coding may transfer strangely. She also suggests that it may be easier to copy the standards from the actual standards website, instead of from this template, but the choice is up to them. As the professional look of the SP will be graded, this tip should help PSTs with the presentation.

#### Other Assignments

In addition to collecting information on the core assignment, I also asked the professors for additional minor assignments that were assessed throughout the course. The purpose was to get a better sense of what else mattered in the course, in addition to the departmental requirements.

Dr. Altair did not provide me with any specific assignment sheets or rubrics, but the way her portfolio was designed, I was able to see some of the assignments as the PSTs made references and links to them in their submissions. As PSTs gave examples for meeting standards, I was able to see that, for example, PSTs were asked to make a concept web for their students, create a WebQuest, and make a video PowerPoint to teach a concept.

Dr. Deneb, on the other hand, gave me information about a Fake News assignment she gave her PSTs. For this task, she provided me with both the task description and the rubric, as well as a sample submission by her student, Rukh. The submission does not have a grade, but I can assume that since she chose to give me this one, it embodies what is expected from this task. I will describe the task and the rubric over the next several paragraphs.

#### Fake News task description and rubric.

Dr. Deneb has a module for her class on Fake News. The module begins with a short introduction, and then contains definitions of what it means to be fake news. As this is a technology-based class, in addition to the prose, Dr. Deneb links to a video about fake news, as well as a virtual training on how to spot it and how to help future students be aware of when the news they are consuming may not be entirely true. I was unable to view the links because they had expired.

Next, Dr. Deneb created a short video and a virtual book about being critical consumers of media and provided two headlines that could be suspect. She tasks her PSTs with using three sources to confirm or contradict the headlines. She also tasks her PSTs with creating their own short videos or virtual book to teach their students about spotting fake news. She includes a red text note that in their work, they need to be creative and use their own images, not just the ones she provided. Dr. Deneb also models citing her sources.

The task has two rubrics, and it is not clear which one is used to assess the PSTs. In both, there are three elements that are the same, that the video or virtual book is age-appropriate for learners (out of two points), that the PST created two or more activities for their students to try spotting fake news (out of two points), and that the artifact created has clear images, as well as correct spelling and grammar (out of one point). In one of the rubrics, however, there is an additional element about correctly determining if the given headlines were real or fake (out of two points). It is possible that I was given two different versions of the task in one, which would explain the double rubric.

# Graded PST submissions

In this section, I analyze the PST submissions for the course assignments. Analyzing what is submitted adds to the understanding of what is being assessed in a course and how. Dr. Altair provided me with six submissions and Dr. Deneb gave me two. Looking across all eight submissions, they all appeared to be relatively similar, although they are from different years. Thus, I decided to not analyze in detail the two submissions from 2015, and only looked at the more recent submissions. I chose to analyze one submission (Okab's) in detail to provide a picture of what the core assessment looks like. For the remaining five submissions, I provide more concise descriptions of my analyses that highlight any major differences from Okab's submission.

As a general overview, I have created this chart (Table 7.1) to keep the submissions and their key similarities and differences clear. I have color coded two of the major trends that I noticed. First, Dr. Deneb's PSTs had a section for their description and understanding of the standard and a section for the artifacts that was kept separate for each standard. Dr. Altair's PSTs, in contrast, embedded the artifacts into the descriptions. Second, I found that secondary PSTs included only one or two artifacts to demonstrate their understanding for each the standards, while the elementary PSTs used any number, varying depending on what they felt represented their understanding.

Table 7.1 Summary of Key Components in PST SP Submissions													
	Instruct -or	# of Artifact	Woven or Separate	Types of Artifacts	Summary Page	Header or Dropdown	Grade level						
Okab	Altair	2	Woven	Own work	Yes	Header	Secondary						
Alshain	Altair	Any	Woven	Own work	Yes	Header	Elementary						
Tarazed	Altair	Any	Woven	Own work	Yes	Dropdown	Elementary						
Суд	Altair	Any	Woven	What learned	Yes	Header	Elementary						
Sadr	Deneb	1	Separate	Own work	Yes	Dropdown	Secondary						
Farawis	Deneb	Any	Separate	What learned	No	Header	Elementary						

### Dr. Altair's SP.

I group my analyses here by instructor, because there are some differences between the two sections. I have chosen Okab's submission as the one to give the full detail, because it was one of the most recent submissions. My theory is that the more recent submissions will be more closely tied to the current task description, and thus more telling about how this task is completed in its current form.

## Okab's SP.

Okab is a PST who is hoping to become a math teacher. His SP opens to a cover page that has a large image of a chalkboard with the word "Portfolio" written on it with fancy handwriting (B11, p. 1). He does not include an image of himself. His first paragraph gives a brief introduction about him and his thoughts on school. From this introduction, we can learn that Okab believes that teachers can be positive influences on students' lives, and it was his own teachers that inspired him to become a teacher.

In his second paragraph, he explains the purpose of the portfolio. He acknowledges that the standards have many sub-standards, but that he will only be addressing some of them. He provides a hyperlink to the full set of standards. This indicates that he has read the task description and can explain the purpose in his own words.

In his third paragraph he introduces his artifacts and explains the grade level and subject area. He also explains why he chose this range for his artifacts, both from a personal and professional standpoint. This demonstrates that he understands the purpose of the task and can explain it in his own words.

Across the top of his website, Okab has a clickable header for the standards. For simplicity, I will go across them from left to right. While some of the other samples numbered the standards

one through seven, Okab instead gives them the names as written in the standards (i.e. Since Standard 1 is about the learner, Okab calls this tab "Learner") (B11, p. 2). As one flips from tab to tab, the header remains at the top. Okab's name also remains at the top, written in black font on a dark wooden background. Each page is designed as a stack of papers sitting on a wooden background.

On each standard's page, Okab puts the number and key word in large font and underlined, justified left. For example, on Standard 1's webpage, Okab starts with "1. Learner" (B11, p. 2). He then copies the standard language directly, although only includes some of the substandards, presumably the ones he is going to connect to. Below the standard, he includes an image (for Standard 1 the image is of a corkboard with the word "knowledge" pinned up), and a link connecting to the source of the image (ibid). Under the image, he uses a paragraph to describe the standard in his own words. For Standard 1, he explains that technology is always changing, and that there is a lot to know, so working together with colleagues is a good way to keep new technology in the classroom. After his description, he then describes and links to two artifacts that demonstrate how he has met the standard.

As every tab starts with the standard copied directly, a cited image, and then the standard written in his own words, I will not mention it for each page. I may only describe the photo to help give a visual for what Okab is attempting to portray.

For Standard 1, Okab provides two examples of how he has demonstrated that he has met it. As his first example, Okab discusses a class task about finding examples of activities that are either good or bad at integrating technology. He links to two images of the discussion post he made about the activities he found. He seems to imply that this discussion forum was a model for collaborating to learn about technology, as required in the standard. As his second example, Okab
talks about when the course started and everyone introduced themselves on the discussion board. He shares a link to a screenshot of his introductory post. He claims that by introducing themselves to each other, the PSTs in this course created a learning network, which is also part of the standard.

Standard 2 is about leadership and empowering students through the use of technology. As his first artifact, Okab links to a position paper that he and his peers wrote in class about Internet filtering. In addition to linking the full paper, he summarizes the main points in a paragraph and explains that too much censorship limits the flow of information, and thus for students to learn properly, they need better Internet access. For his second artifact, he links to his discussion post about an activity he found about launching bottle rockets using a simulator site from NASA and discusses how one can find and share technological resources. He uses an image of a red king in front of a long line of white pawns. I presume the image is about the role of the leader, although his rationale for choosing the image is not given.

The third standard is about digital citizenship. His image is a Wordle including various words about citizenship, society, and rights. As his first example, he included three artifacts from his Moodle Project (which is the course assignment to create an online activity for a class of students, and then to take and review a peer's activity). He first includes the report he wrote about his experience creating and grading the lesson. He then includes the report he gave to a peer about her Moodle project. Third, he includes the report he was given about his project. By showing these three documents, he claims that he had been part of a community of critical friends. For his second example, he includes a link to the class page that he made for class. This page demonstrates an attempt at making a website for students with links to activities for understanding

bivariate data. He claims that the website will help develop his students' "digital literacy" (B11, p. 4).

Standard 4 is about collaboration, and Okab describes this standard as about accepting feedback from peers and students and taking criticism with humility, while the standard talks more about working together to find resources and to consult with experts. As his first example, he shows a sample assignment that he created and explains that since it is challenging for students who are unfamiliar with using the requisite technology, he would have to design a lesson before this task on figuring out the challenges of the technology. His second example is the prep work that went into writing his paper on Internet filters. He shows that by working together in GoogleDocs, his team was able to share ideas and prepare for their paper.

Standard 5 is about using technology to design authentic learning experiences for students. As an image, Okab shows a man sitting at a computer while also writing in a paper notebook. For his first artifact, Okab links to a detailed Webquest that he made for his students about playing with data to measure correlation. The Webquest is filled with comics for entertainment, as well as links and activities for learning to understand how to use data. It also includes a detailed rubric for how students will be graded. The second artifact is an example of another activity he designed where students can choose two types of sports data and then run correlations to find if there are any connections. Both examples use sports as the "authentic" data, which indicates that he is thinking about how to connect math to real life, but also that he still learning about how to expand his connections into multiple different contexts.

Standard 6 is about facilitation and helping students learn to design their own problemsolving process and to "nurture creativity" (B11, p. 7). As his first example, Okab shares the screencast to introduce and handouts to direct students in an interactive activity about using data to

create graphs. The screencast is a six-minute long YouTube video that appears to be a recording of a mini-lecture on data that was taken during a class (as he asks questions and pauses for answers) while the image seems to be what was on the screen during the lecture. The handout then explains the roles and the task for the students. It appears that the creativity component comes from allowing the students to choose between two data sets as they see fit. The second artifact is a very detailed concept web that Okab made to model a lesson. He claims that this web models how to present ideas, provides students with expectations from the lesson, and teaches students about concept webs.

The final standard is about being an "analyst," and suggests that a teacher must be able to make sense of data to make decisions. Okab claims that this is "one of the most important standards" (B11, p. 8). Along with formative and summative assessment data, which is mentioned in the standard description, Okab adds student feedback to the list of data that teachers should be using. His first artifact is excerpts from the Moodle lesson that he created, because as a web-based project that students were to complete at home on their own pace, Okab embedded quizzes that would give instantaneous feedback, as well as methods for students to reach out for help as they needed it. It is not entirely clear how this represents him using data to make decisions. However, as his second artifact, Okab adds that as students progress through the activity, he can see how they are doing and what they have left still to complete. He then used this data to email students to ask how they were doing and offer support. In this way, he is able to use the technology of the Moodle to help him monitor and aid his students.

Okab ends his SP with a conclusion page where he reflects upon his learning in Course BD. He says that he has learned how to use webpages and websites to enhance his learning. He feels like he is confident in his ability to use the technology to enhance his lessons, and his biggest

challenge will be to avoid getting stuck using more traditional teaching methods. To prevent this, he says he will actively work to stay "up-to-date with current technologies" and will work with his future colleagues to share ideas (B11, p. 9).

From this portfolio, one can see that Okab has had experience using many different types of technology, from websites, to screencasts, to Excel. He has had opportunities to give and get feedback from his classmates and has developed a community of learning. He also has reflected upon his experiences, has noted potential weaknesses, and has a plan for continuing to be a technology-infused classroom teacher. From this submission, it is clear that Okab feels comfortable using technology and knows how to use it enhance his future teacher. As there are no teacher comments on this submission, but it is used as one of the samples for other PSTs to follow, it is likely that this portfolio represents what Dr. Altair is looking for in the SP.

## Alshain's SP submission.

On her homepage, Alshain gives two photos. One is a headshot, and the other is a picture of her sitting at a table with a young student. This difference highlights the feel of the portfolio, which is very kid-focused. Her description on her homepage focuses more on her teaching philosophy and her teaching goals, and she does not explain why she wanted to become a teacher.

For each standard page, unlike Okab, her labels are simply, "Standard 1," Standard 2," etc. and does not include a word as a descriptor. Alshain starts with writing the standard in bold, and then including the sub-components as a bulleted, unbolded list. She does not highlight which components of the standard she will address. She then includes an image (with a citation at the bottom of the screen), although there is no explanation for her photo choice. She describes the standard in her own words, using about a paragraph of space. At the bottom of each page, she also includes a list of state student technology standards, both with codes and descriptions. As Alshain describes and explains how she has met each standard, she uses a variety of hyperlinks and examples. For Standard 3, for instance, she provides six artifacts to demonstrate how she has met the standard (B12, p. 4). For Standard 5, however, she only uses one artifact (B12, p. 6). For her other evidence, she just names course assignments that demonstrated her skills. Thus, it appears that Alshain has chosen to pick artifacts based on how well she feels they link to the standard, rather than choosing a preset number of artifacts each time.

In general, while her portfolio holds a similar external structure to Okab's, she is more fluid in her choices. She is less rigid in how many artifacts she presents each time, and adds student standards to the bottom of the page. From her portfolio, however, one can still see that she has had numerous experiences in learning to use technology in the classroom and has spent time reflecting upon her professional growth.

## Tarazed's SP submission.

Tarazed hopes to teach upper elementary. She has only the one photo on her cover page, a headshot, but she uses light colors and fun pictures to accentuate her standards, which demonstrates her ability to accentuate a website with colors and images. When she writes the standards at the top of each webpage, she increases the font and italicizes the sub-components. Like Okab and unlike Alshain, she does not add the extra state standards at the bottom of each page.

For her header, Tarazed uses a more condensed tab. Instead of creating a tab for each standard, she has a dropdown menu for all the standards. She also includes links at the bottom of each page that allows the reader to advance or retreat one standard as they read. This makes the portfolio read more like a webpage.

Like Alshain, Tarazed does not provide a set number of artifacts for each standard, and instead opts to include as many as she feels are necessary. Each page contains numerous hyperlinks to examples, as well as descriptions for how these artifacts demonstrate her meeting of the standard. As she gives her descriptions, she points explicitly to which substandard she is addressing. This explicitness was not seen in Alshain's nor Okab's submission.

Overall, Tarazed demonstrates that she has had many opportunities to develop the required standards, and has spent time reflecting upon her growth. She indicates that she has learned a lot and wants to show it off. From looking at her content, one can learn that Tarazed understands the ITSE standards and knows how to design technology-infused lessons.

# Cyg's SP submission.

There are only a few major differences between Cyg's submission and that of Okab's. First, like Tarazed, Cyg is explicit about which substandard she is elaborating on in her description. She makes clear connections between her artifacts and how it has helped her meet the standard components. Second, like Alshain and Tarazed, Cyg is not bound by a certain number of artifacts for each standard. She appears to use as many or as few as she feels exemplify her learning. Third, Cyg presents her data a bit differently from the others. Many of the times, she embeds the material directly into the webpage itself as a screenshot. Instead of hyperlinking to photos, she just includes the photos with the description. When she does use hyperlinks, while they sometimes are to course artifacts, like her WebQuest, many of her links are actually to resources she has found or learned about. For example, she links to PicMonkey, a site she used to learn to adjust images (B14, p. 1).

Overall, the content of her portfolio is quite similar to the others, but she does show a bit of difference both in how she presents her arguments and how she chooses to link to other sites. Still, she demonstrates the various ways she has grown over the semester and describes how she plans to be a lifelong learner. Overall, it is apparent that Cyg knows how to create technologyinfused lessons and is prepared to use technology in her future teaching career.

### Dr. Deneb's SP.

Dr. Deneb provided me with two PST submissions of the SP. These submissions are provided to the current PSTs when they are preparing to write their SP, and thus they are likely good models of what she expects from the project. One of the samples is for an elementary PST and the other is for secondary. As I have already given full details about Okab's submission and the projects are still similar, for Sadr and Farawis, I will be more concise, as I was with the other Altair submissions. I will primarily note where there are similarities and notable differences.

# Sadr's SP Submission (with respect to Okab).

Sadr is a PST who is focusing on secondary education. His portfolio has a black header bar and a cave wall backdrop. At the top of every page, in large white font (on the back bar) says, in all capital letters, "SADR'S COURSE BD SP" (names adapted for anonymity). Then, also in the back header are links to the different pages in the website, with "home" as the cover page, Standards 1-7, exhibits, and reflection. Not all the links fit across the top, so there is a "More..." button that is a dropdown for the remaining links.

On Sadr's introductory page, there is a spot that appears like it normally has a photo, but it is currently missing. This might be because the sample is not current, or because it was taken down in response to becoming the model submission. He starts with a few sentences about himself, mainly mentioning the grade, subject area, and region in which he would like to teach. He then gives a three-paragraph description of his teaching philosophy, going into more detail than Okab did. To strengthen his philosophy, Sadr cites academic scholars, and then includes a reference list at the bottom of the screen. He then introduces his **SP**, explaining how it connects to his teaching philosophy and will demonstrate how he has grown to meet the current technology standards.

Unlike Okab, Sadr lists the entire standard and sub-standards on each slide. He also does not dedicate a paragraph to explaining what the standard means to him, and instead weaves it into his description of his artifacts. Furthermore, Sadr uses only one artifact for each standard, although he describes it in more detail and with extra commentary. He separates his links from the commentary, opting for an "exhibit" section at the bottom of each page, instead of hyperlinking in as Okab did.

It seems that the differences in submissions may stem from the models and templates provided. The samples for Dr. Altair's class had the hyperlinks woven in, in a more website-style design, while Dr. Deneb provided the paper template that had the components separate. Interestingly, despite Dr. Deneb saying that that PSTs must bold the sub-standards, neither Okab's nor Farawis's submission does this. It is possible that this requirement was added to the task description in a later year from the submissions. Interestingly, while neither task description says how many artifacts are needed, Dr. Altair's PSTs include at least two on every slide, while Okab's includes only one, but in more detail. It is likely that this stems from how the SP was described in Dr. Deneb's class.

# Farawis' SP Submission.

Farawis has a slightly different type of SP submission. She still uses a web format that is divided into pages based on the standards, but she gives her submission a creative theme that aligns with the lessons she created during the semester. Her cover page has numerous photos of her chosen theme, and the header on each page also contains matching images.

One thing that stands out in her submission is her language and grammar. Farawis's comma usage often interferes with a clear understanding of what she is trying to say, and I found that I needed to reread some sentences to understand the meaning. For example, she writes "When I as the teacher, have an understanding for each individual student, and strive to be a caring, loving individual my classroom feels complete" (D7, p. 1). While once parsed this sentence conveys meaning, it demonstrates that either grammar is not her strongest skill, or that she did not spend time editing this submission. Misplaced commas, run-on sentences, and/or strange phrasing are common in her SP submission.

Another thing that stands out is the set-up of the standard web pages. Like Sadr, she copies the full standard with all the sub-standards onto the top of the screen. She then describes in prose what the standard means to her, and then spends a few paragraphs talking about how she grew and developed this semester to meet the standard. When she includes hyperlinks, they are often to sources, not her own work (similar to how Cyg made links). Some of the artifacts she posts are of things she learned, not things she developed. As an example, she embeds a video that she watched in class (D7, p. 1). When she does include her own work, it is normally at the bottom of the screen.

Overall, Farawis's submission is much more explanatory and less demonstrative. To claim she has met the standard, she tends to discuss different modules from the semester and how they relate. When she wants to re-use a model, she includes a hyperlink to the standard where she already discussed it. She also does not include a final page that most of the other submissions had, with a summative reflection.

At first, it is a bit confusing why Farawis's portfolio is used as a sample submission when it does seem so different from the others. Perhaps the purpose is to show that being reflective is as,

or more, important than being able to link to documents created during the semester. What she misses in artifacts, she makes up for in dedication. It is also clear that she used the template, although she did not bold her sub-standards or include many artifacts. Her use of technology is clear and she shows what she learned in a slightly different way. She is descriptive and uses many images. It is, however, quite a contrast.

# Rukh's Fake News submission.

Rukh submitted a word document in response to the task. She first wrote about the two headlines and found three sources for each to verify their veracity. Interestingly, she also added her initial thoughts about one of the headlines, but explained that she double-checked with sources anyway.

She also included a link to a short video that she built. Her video is 2 minutes and 16 seconds long. It is a Mission Impossible themed video about spotting Fake News. There is the Mission Impossible theme song playing throughout and the letters appear on the screen as if they are being typed. The video introduces the students to the definition and provides a list of five strategies for checking the truthfulness of the headline. She then tasks her students with working to determine if two headlines are true, and then provides a group task for figuring out how to spot fake news. Without Dr. Deneb's comments or knowing the target age group for this activity, it is hard to determine how Rukh scored, but since it was given as an example submission, I am assuming that this represents what Dr. Deneb was expecting.

From this submission, one can see that Dr. Deneb values citing sources, as Rukh uses at least six. Dr. Deneb also values creative presentations, as using a theme and adding music was not explicitly asked for in the task description, yet this submission was the sample I was given. It appears that, if this is a representative sample of tasks in Dr. Deneb's class, then Dr. Deneb

frequently asks her PSTs to both demonstrate that they understand the concept themselves (as in the word document) and knows how to use technology to reach students (from the video).

## Interviews with Dr. Altair and Dr. Deneb

# Dr. Altair's interview.

Dr. Altair describes her course as the "methods course for technology for all pre-service teachers" (transcript). In the semester when I collected my data, there were three professors in the department that taught the course. Dr. Altair was teaching the fully online course, and the other two were teaching hybrid classes.

Dr. Altair expects that the PSTs have already taken the curriculum course and assessment course, and now the focus is on integrating technology into the classroom. She expects that the PSTs know how to lesson plan and are ready to apply this process in her course. She also expects that they have already taken an assessment course. She strongly recommends that PSTs do not take Course BD in the same semester that they are doing their student teaching, although it happens occasionally, and those PSTs tend to struggle.

Dr. Altair finds that PSTs often come into the course feeling quite confident about their technology skills. They know how to do the "twittering and the facebooking" but quickly find that knowing how to use technology is the not the same as knowing how to use it as a teacher to enhance learning (transcript). Thus, Dr. Altair teaches them to be reflective about what they know and what they do not know, and expects that by the end of the course her PSTs are confident that they can figure out what they need to know in order to teach with technology.

This class is designed around the ISTE technology standards. PSTs need to know about being digital citizens, understand copyright law, be able to teach their students about digital citizenship, assessing students with technology, and more. When asked what she expects from the

PSTs after taking the course, Dr. Altair said that she expects that the PSTs know "how to routinely integrate" technology into the classroom when it makes sense to teach the content (transcript). She emphasized that technology is to be used to teach the content. In fact, Galaxy University is considering dropping this technology class and integrating the concepts into all the courses, as they would expect the PSTs to do in their own classroom.

The standards for this course come from the department. There is a common assessment, the SP, and it has a common rubric. Other than that, however, it is up to the instructor how to teach the course. Dr. Altair gives an assessment at the beginning of the semester that helps her gauge her class's knowledge and she uses this information to adjust her course. She changes the focus and the level depending on the results. She also makes changes based on the feedback and comments she receives from previous iterations of teaching the course, as well as the notes she takes during the course about how certain activities and rubrics panned out in practice. How she implements the course is fully her choice.

Dr. Altair grades students typically with rubrics, although not always for the written assignments, like the discussions on the forum. She does not have any tests, and instead, PSTs typically demonstrate their skills through projects. The course is designed using a module format, and then throughout, PSTs complete projects to practice the new skills. In her course, PSTs write a "technology infused lesson," design a web-based lesson, create a WebQuest, work in teams to learn about a technology and present it to the class through a video, and more. In the SP, PSTs are required to look across the entire course, reviewing everything they have written and designed, to bring it together in a final portfolio. While I was told about each of these assignment in general, I did not get specific information on grading, point values, or weighting for anything other than the SP. She did say that the tasks and rubrics are the same for each assignment given in the different sections taught by her, but there is no guarantee that sections taught by other instructors are include the same tasks.

I asked Dr. Altair to qualitatively describe the different grade levels (A, B, and C) for this course. Dr. Altair thought about is briefly, and then said that if a PST meets all the expectations and standards, she gives them an A. If one goes beyond and exceeds expectations, they can get the A+. This indicates that the target grade, the A, is about doing what was expected. PSTs who receive a B in the course are just not quite "getting it" and have some things missing or incomplete or unclear (transcript). PSTs who receive a C have done the bare minimum. In this class, PSTs need a C to pass, and so PSTs who receive a C demonstrate that they know what is happening, but just are not quite showing it or doing it at a level that she thinks they should. She feels confident that the PSTs who receive As and Bs can enact what they learned in the classroom, but with a C, she is not sure there is the transfer from knowing to doing. A student fails the course if they get below a C in either the course or the SP. For the SP, to get a C, one needs to be meeting acceptable in all areas and target in a few. To get a B, most areas are at target. It was not clear if a PST could get a C if they were on target for half the standards and unacceptable for the other half, such that they still had the necessary 14 points, but it was clear that they needed to meet target at least somewhere.

Dr. Altair said that this concept of the rubric and points is fairly consistent across all her assignments. She tries to describe clearly in the rubric what target looks like as well as the lower levels, and then the PSTs who get the full points get the A, and it descends from there. We had discussed her giving me a de-identified sample of a C work, but that never happened.

I asked Dr. Altair about potential alignment between her sections and others, and she responded that lack of alignment is a current issue. She said there have been discussions about how to balance professional academic freedom with ensuring that all PSTs get the same quality

experience (transcript). She said that she talks quite a bit with one of the other two instructors, and the hope is that by discussing, the sections become more similar. Even the SP, however, which has the same task and rubric for all sections, is still graded by the individual course instructors. She said that for the "most part" using the same rubric unifies the SP across the sections, but Dr. Altair also said that there is subjectivity involved and that she knows that at least she has added some components to the SP and grades for them, and that is not the same in the all the sections. Some of the differences in the SP stem from the different assignments that occur during the course of the semester, since the SP is a culminating project. In the end, Dr. Altair thinks it is likely that a PST taking a different section may get a different letter grade than if they had taken her course, and she feels that this is problematic. However, despite the variation in grades, she feels that the course still teaches the same general material regardless of the instructor and that the claims from the course can still be made.

# Dr. Deneb's interview.

According to Dr. Deneb, the purpose of Course BD is to prepare PSTs to use technology to support teaching and learning. She wants PSTs leaving her course to know about technology and to know when it may be appropriate to use technology for enhancing student learning. She wants them to inquire about the applicability of a technology before using it, and to examine it before implementing. She also wants PSTs to act professionally and collaboratively. When they leave the course, she wants them to be able to do all of the above. She acknowledges that they may be lacking the confidence and time to actually use the technology, but at least they will have the requisite knowledge. As she said, when leaving the course, she hopes that the PSTs will have a "treasure box" of lessons that they can do, and will not stick to only one type (interview notes). The PSTs in Dr. Deneb's section are generally juniors and seniors who have been admitted to the College of Education. The PSTs range in age from 20-60, and are primarily white and female. They have already taken a curriculum course, and are expected to have a basic knowledge of technology, such as how to turn on a computer, use a search engine, and have familiarity with the Office Suite or equivalent. She then takes this basic knowledge of technology and transforms it into how to teach using technology.

Dr. Deneb designs her course in an order that builds around what she expects a teacher to need to know first when starting a school year and to align to the ISTE standards. She begins with ethics about technology, how to communicate with parents and students, and then moves on. She also builds up the complexity of the technology, starting with building a website, and then a WebQuest, and then a Moodle. There are two assignments that she must assign, the SP and the moodle project, as they are both commonly agreed upon by the department. The moodle task is to design an online learning management system. Interestingly, while Dr. Altair also mentioned the moodle task, she did not say that it was a required task. Other than the required two tasks, Dr. Deneb uses her own discretion on what to teach and what to include. As she explained, "as long we help them achieve the standards, we're good" (interview notes). To help her decide what to include, Dr. Deneb attends an annual conference attended by current classroom teachers. She sees what they are doing in the classroom and then decides what needs to be included in her course. In the semester than I collected data, Dr. Deneb was having her students do several assignments, including making a teacher webpage, creating an online poll, completing an "hour of code," solving a problem using a spreadsheet software, and designing a lesson on spotting fake news.

In our discussion, we talked a bit about the grading for this course. Dr. Deneb uses a point system, and each assignment has a rubric with the points attached. Dr. Deneb said that the moodle

and the SP are weighted the most, as they are required assignments, and because if you do not pass the SP, you cannot pass the course (as it is the core assignment). The other assignments are worth up to fifteen points each, with 151 points possible (interview notes). Because she uses a course site, PSTs are able to see their grades throughout the semester, and are able to see what they have earned so far and what they can still earn. Dr. Deneb uses rubrics for all her assignments because she finds it to be "objective"; PSTs are able to see "the points they will be awarded along with the criteria that will be assessed" (ibid).

I asked Dr. Deneb if any course components were not assessed. She said that online discussions used to not be assessed, but she had to change that because otherwise the PSTs would not participate. Now, she requires that they do things like initiate a thread, respond to two classmates, or find a resource that aligns to their thinking and they earn points for doing so. In-class discussions are still not graded, but they stem from the online discussions.

To get an A in Dr. Deneb's section one must meet several criteria. First, they must earn a majority of the points on all the assignments. Second, they must act professional. This includes coming to class prepared, being engaged in the discussion, turning in assignments on time, and taking the lead in class when appropriate. Third, to get an A, a PST must show a willingness to develop their professional self as a teacher who can integrate technology in the classroom. One way to do this is to be "willing to participate in conferences" (interview notes).

PSTs lose points if they turn in work late. If they do poorly on an assignment, as in less than a C, they may ask for permission to redo it. PSTs can fail the course for various reasons. Because of the point values, doing well on everything, but not doing the moodle, for example, can still lead to failing.

I asked Dr. Deneb about the alignment between her sections with other instructors teaching the same course with respect to grading. She said that she has spent time talking with Dr. Altair, so she feels that their courses are pretty similar. Otherwise, however, she says, "people are pretty private about it" (interview notes). Dr. Deneb spoke about academic freedom, and stated that as long as the professors were doing what they needed to do, alignment did not need to be perfect. She explained that the SP rubric is very general since all the instructors use it. Dr. Deneb also said that the moodle project may be required in all sections, but professors use individual rubrics to grade it. Essentially, the instructors are all looking for similar things, but the exact grading will vary. She said she was unconcerned about small grading differences because that it not what matters, and while it is possible that a PST could fail one section and pass had they been in another, it is not probable. Essentially, Dr. Deneb said that she did not want to be a "robot" and instead preferred teaching a course that was "representing me and what I feel my students need" (ibid). Thus, her course is her own, but she still adheres to the standards and ensures that her PSTs learn what they need to learn.

### Looking Across

I spent time analyzing across the different indicators to try to get a sense of what really matters in this course. Unlike for Course A, I only really have rubrics and weighting for the core assignment, but as it is worth about 50% of the grade in both sections, it still allows me to get a sense of what is graded and how. I built matrices for both instructors, looking at what is claimed in the syllabus as course objectives and matching it with the core assignment. These charts do not include the standards that were discussed in the interview, but only from the syllabus.

Table 7.2 Matching SP Rubric Elements with Dr. Altair's Syllabus Objectives											
			Course Objectives/Standards								
Weight			1	2	3	4	5	6	7	8	
Core Assignment Rubric Elements	Α	0.13	1								
	В	0.13		1							
	С	0.13			1						
	D	0.13				1					
	Ε	0.13					1				
	F	0.13						1			
	G	0.13							1		
	Professionalism	0.09									
	Totals		13%	13%	13%	13%	13%	13%	13%	0%	

Table 7.3 Matching SP Rubric Elements with Dr. Deneb's Syllabus Objectives										
			Course Objectives/Standards							
Weight		1	2	3	4	5	6	7		
Core Assignment Rubric Elements	Α	0.13	1							
	В	0.13		1						
	С	0.13			1					
	D	0.13				1				
	Ε	0.13					1			
	F	0.13						1		
	G	0.13							1	
	Professionalism	0.09								
	Totals		13%	13%	13%	13%	13%	13%	13%	

Looking across the two charts, it seems that the course is fairly balanced in assessing all the standards listed in the syllabus. This is in part because the SP rubric is designed specifically to assess each of the standards. Dr. Altair has an additional standard in her syllabus, but it is not assessed in the core assignment. In terms of alignment, it appears that both instructors grade the core assignment along the exact same standards. What it means to score well within each standard, however, may be more subjective, but that cannot be ascertained at this time.

If one adds up the total percentages across the bottom on the chart, one will see that it only adds to 91%. This is because there is an additional 9% of the SP rubric that pertains to professionalism. Professionalism is not one of the course's explicit standards, but it still retains a portion of the overall grade. This professionalism grade was also seen in how late assignments got points deducted, as that demonstrated that turning in assignments on time was worth part of the grade.

Interestingly, the two professors appear to have different qualitative descriptions of what it means to get an A in the course. While Dr. Altair said that meeting the expectations was an A, Dr. Deneb added the professionalism and willingness to learn more as components of an A. While it might seem like this would lead to students scoring higher in Dr. Altair's class, the opposite might actually be true. According to Dr. Altair, Dr. Deneb once looked at her rubrics and remarked that Dr. Altair was holding the PSTs to a higher standard than she was for everyday assignments. Thus, while Dr. Altair's qualitative description is less involved, her rubrics may require PSTs to be more rigorous in their assignments.

As I did not receive information on how the rest of the course was broken down specifically, I am unable to make any claims about how the full course is weighted with respect to the rubrics. The only rubric I did see (for the Fake News assignment) was worth only 7 points, so it was not enough from which to build any conjectures.

There was a bit of tension between the sections of the courses. Dr. Altair seemed to be interested in increasing alignment and making sure that how PSTs were graded in her sections was the same as the in the courses taught by other instructors. Dr. Deneb, however, believed more in academic freedom and expressed trust in both her and her colleagues experience to be able to

teach the PSTs well without conforming to a single curriculum. They both cared about the experiences of their PSTs, but in different ways.

# Summary

All in all, it appears that in this course PSTs get several opportunities to learn about new technologies and to try writing lessons in ways that incorporate these technologies into the classroom. Both professors mentioned that it was important that PSTs learned not just about technology, but when it was appropriate to use it for a lesson. They explained that the purpose of teaching was still the content, and that technology was to be a means to make this content more accessible. By the end of the course, Drs. Altair and Deneb want the PSTs to feel comfortable not only using what they learned, but also in looking into new and changing technologies and be curious learners about what comes out next. This course is not the end, but a beginning.

"I want you to be confident when you are in a classroom, that even if you don't know what to do, you know how to find it and figure it out." – Dr. Altair (transcript) SECTION 3: ANALYSIS AND IMPLICATIONS

#### Chapter 8: Looking Across the Cases, an Analysis

By collecting the syllabi, the core assignments and rubrics, and the PST submissions, as well as by interviewing the instructors, I was able to learn about each of my cases. Looking at the data within each course afforded me the opportunity to learn what was being taught and how it was assessed, as well as start to form an understanding about the claims that can be inferred about a PST passing each course. I was able to get an understanding of each course individually and I shared those findings in the prior chapters.

In the concluding chapters of my dissertation, I look both at what one can learn about GU from the collected data and about what tensions arose about assessment as I analyzed. In the next chapter I will look closer at the tensions, but in this chapter, I look at the broad questions for understanding what happens in the middle of a teacher preparation program like GU.

I collected and analyzed my data not only to know what was happening within individual cases, but also to get a broader understanding of how PSTs are taught, what they are taught, and how they are assessed in the middle of their teacher education program. This is a qualitative project, so my results and findings will not be generalizable to all teacher preparation programs, or even all similar teacher preparation programs, yet my hope is that by looking across the three courses, I will be able to understand the tensions and questions that arise when considering middle-program courses which may be broadly relevant to all teacher preparation programs, and not just the one at GU. I make the claims in this chapter with the caveat that in order to make claims about all teacher preparation programs, much more research would need to be conducted, and the sample size would need to be considerably larger. However, because I chose a typical case with specific core courses, I hope that this research will set a foundation for how one can look and determine how we are preparing teachers on a larger scale, and about what themes exist. As

discussed in the literature review, assessment is a big category with a long and complicated history, and how we assess in teacher education is not uniform or constant. Nevertheless, by looking closely at these few cases I aim to highlight what can be learned about a specific university when the methods I used were completed.

I arrange this chapter around questions. Some of the questions were those I had before collecting the data and was fortunate to be able to answer with the information I gathered. Other questions emerged from the data itself, as I spent time reading and rereading my sources, coding the information, and talking with my advisor. These questions (and answers) presented themselves, and so I include them, as well. Some of the material presented in this chapter cannot be found in the course chapters. This is because as I considered the questions, I referred not just to my own course analyses, but also to the original documents and some of the interview data.

The questions discussed in this chapter are:

- What can we learn about a course from different course materials?
- What is the purpose or objective of a course, according to the individual documents?
- How do teacher education courses vary by instructor and over time?
- How does the order of learning affect what is taught and graded?
- What does a grade in the course tell us?
- What the purpose for assessing dispositions, especially professionalism and rule following?

# What can we learn about a course from different course materials?

As I reviewed my data, it was clear that what one understands about a course changes based on which document or source is consulted. The National Council on Teacher Quality conducted a study in 2006 looking at course syllabi and required texts to understand what PSTs were learning in their courses. While their sample size was substantial and their findings were intriguing, this study lead to questions about if one could really understand what is happening in a course by looking only at these two sources.

Gorski (2009) attempted to understand what was happening in teacher preparation multicultural education courses by looking at course syllabi. He acknowledged, however, that what is put on paper does not necessarily translate to what is really happening in the classroom. I was able to verify this hypothesis as I compared the different claims presented in the different source documents and interviews.

### What can we learn from a syllabus?

Looking at the four syllabi, I wondered what I could learn about what mattered in a course from considering only a syllabus. I found that each syllabus listed the official objectives for the course. In Course C, this was presented as a general concept about what it means to be a teacher. It was then accentuated with a bulleted list of fifteen items that were divided into five major themes. In Course A, these objectives were listed in bold as "essential outcomes" (A9, p. 1). In Course BD, the objectives were copied from the ISTE standards, and for Dr. Altair, there was an additional state standard. Thus, as each syllabus presented the objectives, albeit in different forms, it was still clear that if one wanted to know the official objectives for a course, one could find this from the syllabus.

Each syllabus also provided a bit of information about the assignments that would be completed during the course. For Course C, Dr. Aldebaran did not give specifics about the assignments, but listed the various ways that the PSTs would learn. She described these methods in her syllabus as "Instructional Methods and Activities" (C3, p. 2). There were no weights or due

dates, but PSTs could get a sense of how they were going to develop their knowledge. For Course C, Dr. Polaris included a short note about required, non-graded assignments (such as "readings, short written assignments"), as well as providing a weight breakdown of all graded assignments (midterms, projects, etc.) (A9, pp. 2-3). For Course BD, Dr. Altair listed the general ideas about each of the assignments in the course, and Dr. Deneb provided a calendar for each of the module assignments. Thus, how the assignments were presented and described varied in detail and style, but each syllabus did indicate the myriad ways the PSTs could be expected to develop and demonstrate their learning.

The syllabi also all included something about the core assignment. A core assignment is something that GU requires in all its teacher education courses, so it made sense that reference to it would be included in the syllabus. The specifics of the core assignment were not necessarily given, but enough was presented to get the general idea. Dr. Aldebaran explained how the core assignment was divided into two projects and listed the many ways a PST could fail it. Dr. Polaris did not give any specifics about his core assignment, but stated that it would be worth 50% of the course grade (A9, p. 3). Interestingly, Dr. Altair described the portfolio at least twice, but did not mention that it was the core assignment. Dr. Deneb also did not explicitly make mention of the core assignment, but did include it in the list of due dates. Course BD had an online course site, however, where the PSTs were directed for more assignment information, so the core assignment was likely described in detail there. Thus, from the syllabus, PSTs learned a bit of information about their core, required project.

What I found interesting about all the syllabi was that there was no clear match between the course objectives and the assignments and assessments. If a syllabus reader wanted to know how the **PSTs** would be expected to demonstrate their learning in general, the details were all there, but

to get the specifics, additional information was needed. In none of the courses could I clearly match the assignments with the course objectives. Thus, the syllabus was not the place to see how **PSTs** would demonstrate their knowledge of the individual objectives, nor to understand how each objective was meant to be assessed.

Lastly, the syllabus was replete with information about professionalism and rule following. To measure this, I looked at word count. Whenever the syllabus talked about following directions, using specific formatting, attending class on time, avoiding plagiarism, classroom behavior, I highlighted it and counted it. I then found the percentages of text relative to the full syllabus. Using word count is not an exact measure, and thus I acknowledge that all the results of this method presented here are approximations. Nevertheless, they make a point. 42% of the syllabus for Course C focused on professionalism and rule following, 31% for Course A, 31% for Dr. Altair's section of Course BD, and 19% for Dr. Deneb's section of Course BD . Therefore, across the syllabi, approximately one-fifth to two-fifths of the syllabus was dedicated to this subject. Thus, the syllabi were clear places to learn about how PSTs were expected to act in the course.

In summary, one can learn at least four things from the syllabus of a course in the middle of a teacher education program. One can learn about the course objectives, basic information about course assignments, information about the core assignment, and information about the required professional behaviors for participating in the course. The exact details for these four items will vary, especially in the degree of detail, but they are each mentioned in each syllabus, especially professionalism. If one wants to know how the assessments align to the objectives, however, one will need to look elsewhere.

#### What can we learn from a core assignment task description and rubric?

Looking at my four courses, I also wanted to know what could be learned from looking solely at the core assignment task description and rubric. As a core assignment is a key component of every teacher education course at GU, I figured that looking to see what I could learn about the course from just these documents would be interesting.

One thing that is unsurprising about the task descriptions and rubrics is that they tell the **PST** what must be completed in order to complete the task. Thus, while the presentation of the directions may vary, one can expect to find instructions in the description. What does this teach us, then? It teaches us that the task description and rubric are a source of information on what must be included in the assignment. Interestingly, the purpose of the task is not always included in these documents. Of the four task descriptions and rubrics that I collected, only one, that of Dr. Altair, stated the purpose of the assignment. It can then be concluded that either the purpose is never stated, or that it is mentioned in class and therefore dropped from the task description. If someone were to conduct a study on the purpose of the core assignments, looking at the provided task descriptions and rubrics would not be the source to consider.

All three task descriptions worked in different ways to describe what the PST needed to do for the task. Course C embedded the task description with the rubric, which afforded the PSTs the opportunity to understand weighting and to make their components the appropriate lengths with the necessary details. Also, the components were occasionally broken down into subcomponents where even more detail was provided. Course A gave a very detailed prose description of what should be included and where in the task. In addition to describing what needed to be done, Dr. Polaris gave suggested page lengths to help guide the PSTs in the work. Course BD provided the PSTs with not only the task description, but also a link to a template and sample submissions to provide a working framework. Thus, looking at all three courses, I could clearly see what was expected to be included in each task.

I was curious if looking at these documents would provide me with a clear understanding of how the core assignment would be assessed. I found that there was a lot of variability and subjectivity in the rubric that would make it challenging for someone else to grade the assignments in the exact same way. Nevertheless, the rubrics were designed in such a way that suggested that the course instructor did have a plan for grading and would know what they were looking for. In Course C, Dr. Aldebaran used a rubric that gave a point breakdown of all the components, and the comment that meeting these minimally would constitute a C in the class. A component that was worth 7 points, however, did not provide any explanation of how those seven points would be allotted for partial credit. It was also not clear that, if this was what one needed to get a C, if 7/7would be a C or an A. (I will speak more about anchoring in the question "What does a grade in the course tell us?") Similarly, in Course A, Dr. Polaris listed the components that needed to be completed for 100%, and then said that points would be deducted accordingly. How they would actually be deducted, however, was not clear. There was also not a breakdown of weighting, and when a component had multiple sub-components, it was not written how the sub-components related to each other in terms of weight. For Course BD, the rubric was more filled out, with three levels all with descriptors. The difference between the levels, however, still occasionally was dependent on subjective measures, such as the difference between "exhibits and articulates an understanding" for two points versus "exhibits broad understanding" for three points. For all three courses, it was likely that the instructor understood the grading scheme and could enact it systematically, but as an outsider, it was not clear how the task would actually be graded and

awarded points. Thus, looking at these documents gave me a general, but not complete, understanding of how PSTs would be assessed.

In summary, I found that from looking at the task description and rubric, one could ascertain what needed to be included in the assignment submission and a general understanding of how it could be graded. It was not uniformly the case that the purpose for the task, nor an exact grading scheme, would be included, although it was possible to make some general inferences based on the documents.

#### What can we learn from PST submissions?

One of the data sources that I really enjoyed was looking at the PST submissions to the assignments. Having submissions, especially those that embodied what the instructor was looking for, provided me the opportunity to really see what the task looked like in practice. When looking at a task description, one can imagine what an A submission would look like, but without being in the course, this understanding is limited. As an example, I would not have fully realized how much the order of learning mattered in what was expected and graded. I had not previously fully understood what it would mean to create a unit assessment without taking a methods course, and how it would be graded accordingly. Thus, having the sample helped to demonstrate what could be expected of the PSTs at GU at this point in their learning trajectory.

The submissions also helped to explain the rubric. Especially for Course C, which had the comment about minimally meeting requirements, seeing a graded submission with the corresponding rubric helped clarify the expectations. It turned out that a 7/7 would be an A, and that minimally demonstrating the components would be something like a 5/7. Looking at the submissions also helped with understanding what it meant to need to use "intelligent speculation" when making accommodations for the assessment in Course A, since the prose description of what

needed to be included was highly dependent on having actually been in class when accommodation was covered (A1, p. 4). Thus, looking at sample **PST** submissions helped to bring the task description and rubric to life, by giving a good example of what success looked like in practice. In future studies, I would recommend looking at submissions from just above and below the pass line, as that would give even more information about what really matters and what counts as "enough."

#### What can we learn from interviewing instructors?

This question, what we can learn from interviewing the instructors, is not as clear as the previous two questions. As an interview is dictated by the interview protocol and the conversations between the researcher and the instructor, what one can learn is highly dependent upon what one asks. Thus, the answer to this question will be influenced by the questions that I prepared for the interviews.

One thing I learned from all the instructors is how they qualitatively described what it means to get an A or a C in the course. I chose these letters because an A represented the top grade, and a C was the lowest one could get and still receive credit for the course. Each professor had a different understanding of these grades, but they did have an understanding that aligned to how they graded. I will break down their grading in the grading question.

Another thing I learned was about their opinions on academic freedom versus alignment, as well as how they made choices to include or exclude material from the curriculum or core assignment. I wanted to know which components of the course were theirs to make and which came from the department. I was able to learn that while the department provided the course description and the core assignment, most everything else was individualized. How the daily lessons went, what the texts or other assignments were, how the core assignment was explained and

weighted, all of these were at the disposal of the individual instructor. Thus, what I learned was that at GU, at least in the teacher preparation program, sections of the courses had common themes but the implementations varied. More exact descriptions of the variation are discussed under the variation question.

Overall, interviewing the instructors offered an insider look at the choices, decisions, and opinions of those teaching these mid-program courses. They rounded out what I could learn about the course by providing a human face to the course. Without the interview, I also would not have had the information in order to make the alignment matrices in the course chapters. It was these conversations that provided me the missing link.

# What is the purpose or objective of the course, according to the individual documents?

While the previous section dealt with what could be learned in general from considering the different artifacts, this question focuses more intently on the purpose of a course. To answer this question, I looked at the explicit and implied objectives as given from each artifact.

### What is the purpose of the course according to the syllabus?

According to the syllabi, there were two purposes for each course. The first purpose was to meet the course objectives. The objectives were always stated clearly and near the beginning. In Course C, this objective was to become "caring professional educators for a diverse and democratic society," and this objective was divided into five more specific themes (C3, p. 1). In Course A, there were five essential outcomes, which were to be able to understand principles of assessment, critique and evaluate assessments, construct quality assessments, analyze and use data effectively, and advance professionalism. In Course BD, there were the seven ITSE standards given. In Dr. Altair's section, there was also this added prose purpose, "to critically and creatively apply the concepts, principles, hardware and software associated with the infusion of technology in solving educational problems and meeting challenges in your role as a classroom teacher" (B1, p. 1). Thus, the purpose of the course, according to the syllabus, was to meet the course objectives.

There was also a second purpose, which was to conduct oneself in a professional manner. In Course C, this was explained in the sections of the syllabus that talked about using appropriate vocabulary and grammar, how to engage with technology, and how to deal with late work. In Course A, this was explained with a focus on in-class behavior and how to respond to being tardy. In Course BD, this was explained with how to communicate with peers and how to submit work. Thus, the purpose of the courses was not just to learn the course material and meet the objectives, but also to be able to do so while engaging in professional behavior as described by the instructors.

# What is the purpose of the course according to the core assignment description and rubric?

According to the core assignment task descriptions and rubrics, PSTs needed to be able to demonstrate that they had met the course objectives while adhering to specific formats and styles. Grades were awarded not only for the content, but also for the presentation.

In Course C, PSTs needed to design and develop a unit plan that would teach future students about a chosen topic, while following a plan format that was provided in nearly template form. Thus, PSTs needed to show that they could follow directions, as well as plan individual units, use student standards, and assess student learning.

In Course A, PSTs needed to develop a summative assessment plan. On the task description, it was stated that this assignment would measure the PSTs' ability to meet three of the five course objectives (understand assessment, construct quality tests, and advance

professionalism). Here, rule following and professionalism was not only an implied purpose, it was stated.

In Course BD, PSTs needed to create a digital, summary portfolio that demonstrated how they met the course objectives. Both sections provided the PSTs with rubrics to follow. With this assignment, the purpose then was to be able to show both that they could use technology and that they had mastered the standards.

Looking across the three courses, the common themes were to be able to demonstrate mastery of course standards and to follow directions.

#### What is the purpose of the course according to the graded submissions?

From considering the graded submissions, it appeared that following the required format and using clear and proper grammar was important. Also, PSTs were required to be reflective in their writing.

In Course C, Dr. Aldebaran added comments to the PPp1 submissions for the PSTs to use and change before submitting PPp2. These comments highlighted spelling and grammar issues, pointed out places that PSTs could add to their submission to be clearer and more cohesive, and added suggestions for how to strengthen and improve the submission. Thus, it appeared that the grading, at least the points deducted, came from failing to properly adhere to the instructions or by not thinking enough about the project.

In Course A, comments on the submissions pointed out both when areas were good and when they could be improved. For improvements, Dr. Polaris made comments encouraging the PSTs to "say more" about something or provide more detail. Dr. Polaris made positive comments when he found that a PST was giving a good description or provided a good chart. Thus, it appeared that what mattered was to clearly and thoroughly adhere to all assignment questions.

In Course BD, while I did not get to see any comments on the documents, I was still able to see that what appeared to matter was that PSTs were reflective, that they used technology well, and that they included all the required elements.

Looking across all three courses, I found that the purpose of the course, according to the PST submissions, was to demonstrate mastery of the course standards and to follow directions thoroughly. The purpose here was similar to what I found from the core assignment descriptions and rubrics.

#### What is the purpose of the course according to the instructor interviews?

As I met with and interviewed the course instructors, I asked them to tell me the purpose of the course in their own words. Dr. Aldebaran spoke to me about teaching the PSTs about curriculum methods and instruction, and designing all course readings and activities around how to complete the core assignment. Dr. Polaris said that his course was an introduction to classroom, academic assessment and evaluation. He believed that assessment, if done right, was a good and helpful thing and therefore wanted to help his PSTs know how to design and use assessments effectively. Dr. Altair spoke about wanting the PSTs to be able to effectively integrate technology into their future teaching. Lastly, Dr. Deneb stated that she wanted to prepare PSTs to use technology to enhance teaching and learning.

What I understood from these four replies was that these were the motivations behind the purposes for the courses. Each instructor had a personal connection to the material and wanted to provide a venue where the PST could learn, practice, and grow. Thus, the purpose of the courses, according to the instructors, was to get a chance to develop important skills that would be instrumental in guiding the PSTs to becoming successful teachers.

# How do teacher education courses vary?

As I examined my data, I found that there were many ways in which the sections of the courses I examined had space to vary and change. Going into my research, I held the following philosophy:

Pre-service teachers and instructors are individuals and they should be able to express their ideas in their own ways. Individuality gives PSTs and instructors the ability to construct meaning together, and build knowledge based on shared experiences. It allows the PSTs to express their learning in different ways, matching to their comfort and future plans. Conformity ensures that everyone learns the same thing and makes sure that differences in background won't act as a barrier to the future. It acts as a form of equity as it cares about equal outcomes. Individuality pushes back, saying that this cookie-cutter form of education doesn't allow for the kind of diversity that the future needs, and worries that instead of preparing PSTs for their future, it prepares them for the status quo. Conformists counter saying that unless you provide equal opportunities, you cannot ensure that the status quo won't be so stratified. Thus, it is complicated to decide which is better and how to navigate between the two. (Ellis, **R**., personal notes, September 19, 2017).

Thus, as I looked at my data, I wanted to look for areas of individuality and conformity to see how the instructors at GU reflected (or not) my conflicting philosophy. I found that variability presented itself in a few forms, by instructor and by semester. Both of these forms of variability were reflected in the instructors' understanding of academic freedom.

All of the instructors that I interviewed had conceptions of what it means to have academic freedom. As Dr. Aldebaran put it, she was given a core assignment with which to assess her PSTs, but the focus of this assignment was up to her. Also, she was given a curriculum guide, but how she

chose to implement it was up to her. Dr. Polaris said that he had "freedom within the limit of the course description" (interview). However, he said that if there were other instructors teaching the same course, he was expected to meet with them and collaborate to ensure that they did not have "drastically different versions" of the same course (ibid). Thus, the course was technically his to teach as he wished, as long as there was some sense of consistency. Dr. Deneb said, "as long as we help [the PSTs] achieve the standard, we're good" (transcript). She liked that the course reflected who she was and that she had the freedom to meet her students' needs. Other than the agreed upon assignments, she enjoyed using her expertise to individualize her sections. Lastly, Dr. Altair said that the standards come from the group and the implementation is individual choice (interview notes). However, she felt that it would be better if all the sections were more aligned. Of all the instructors that I interviewed, Dr. Altair was the only one who felt that academic freedom might lead to drastically different courses and material learned, and this could be problematic both for the program and for the PSTs. Overall, all instructors suggested that as long as they met the course description and gave the core assignment, the rest was up to them. Thus, I was unsurprised to find variability across sections and over time. In the next few pages, I will discuss the variations I found from my data.

## How do courses vary by instructor?

In this question, I look at how the course instructors individualized their sections as compared to other instructors of the same course. As I did not interview other instructors for Course C and Course A, the information I have comes from what was stated by the instructors during the interview, as well as from my own analyses.

For Course C, Dr. Aldebaran discussed with me some of the ways that she personalized her course. While the PP is a departmental assignment, it was Dr. Aldebaran's choice to split the
project into two components. She found that by creating a PPp1 and PPp2, her PSTs were better able to create a unit plan. Interestingly, this then also altered how she ended up grading the PP. Dr. Aldebaran chose to make her students get a C or better on both parts individually, not as a combined total. This then differed from how the assignment would be graded in another section. Dr. Aldebaran's PSTs could not have a high PPp1 grade carry them through a low PPp2 grade, but in another section, with a combined grade, it is possible that this could happen. Additionally, Dr. Aldebaran said that she got to control the weighting in her rubric for the core assignment. She said that one semester she helped another instructor with grading their PP submissions, and she had to use their rubric to score it, because otherwise the emphasis might not be in the right place. Therefore, while it might be the same core assignment in theory, it definitely varied by instructor.

For Course A, I had less data to review because Dr. Polaris was the only instructor teaching the course in the semester I collected data, and he tends to be the primary instructor. He did say, however, that when there are others teaching the course, there are slight differences with both the syllabi and the grade weighting. He surmised that this might lead to a half-step grade difference between sections (B versus B-), but nothing too serious. Dr. Polaris also said that because of his expertise with this course, new instructors of the course tend to defer to him. Therefore, while there might be some variability due to instructor, this variability was minimal.

For Course BD, I had concrete evidence of variability between sections because I interviewed two different instructors for the same course. The first major difference was that one section was a fully online course and the other was a hybrid course. If this difference causes variability is up for debate, as mentioned in the literature review, but it is nevertheless a difference in implementation. Another difference was that the core assignment appeared to be implemented and designed a bit differently. For example, Dr. Altair had her PSTs weave their artifacts in with

their descriptions of how they met each standard, while Dr. Deneb had the PSTs create separate sections for descriptions and artifacts. Perhaps the largest difference, however, was the cut score for passing the core assignment. Dr. Altair required PSTs to get earn 14.5 points while Dr. Deneb required 16 points. Interestingly, Dr. Altair said that the PSTs needed to get the 14.5 points on the proficiencies and to equal the 70% as she states, it means that the professionalism grade on the rubric does not actually factor in to the score. I did not have enough information to determine if Dr. Deneb also skips the professionalism. Either way, this shows that despite having a common rubric, Course BD still determined the pass line for the core assignment individually.

#### How do courses by the same instructor vary over time?

As different as sections can be when they are taught by different instructors, maintaining the same instructor also does not guarantee that the course remains constant. In addition to asking the instructors about what makes their sections unique, I also looked into the data to determine what changes an instructor might make when re-teaching the same course. Two general themes emerged when I looked at how the courses could change when taught by the same instructor. The first theme was about adapting the course to the PSTs enrolled and the second was about updating the course as new research and ideas come into the field.

Dr. Altair was perhaps the most explicit in explaining how she adapted the course each semester to her enrolled PSTs. She gives a pre-assessment at the beginning of each semester and uses the results to influence how she runs the course. If PSTs are particularly strong or weak in a certain area, she changes how she teaches to meet their needs. She also said that there have been times when she finds that she gives a task or assignment and nearly everyone gets the task wrong. She then uses this information to reteach the material and assign a new task. Thus, she uses how the PSTs are learning to adjust her curriculum in real time. In addition to adapting to her current PSTs, Dr. Altair also mentioned that she has made deliberate changes from semester to semester. She takes notes during the semester and adjusts the curriculum moving forward based on what worked and what did not. She also uses the feedback that the PSTs provide at the end of the semester to make changes. Thus, if once were to take Course BD with her two semesters in a row, the course would not necessarily be the same. PST experience, previous results, and current learning speed would all influence how she taught the course.

Slightly similar, but not quite, Dr. Aldebaran also changes up her course from semester to semester. Sometimes these changes are the reaction to the enrolled PSTs and their interests, but she also makes changes based on her own shifting interests. She acknowledged that teaching the same course over and over could be boring, and so she makes small changes to keep the course interesting for her, as well as relevant for her students. One of the examples she gave was with the focus of the core assignment. One semester she required that the unit be based on a local historic site, and one semester it was based on a community action project. Such changes keep the core assignment in tact, with the goal of writing a unit staying the same, but shift the focus of how the unit is planned and designed. Thus, taking Dr. Aldebaran's course two semesters in a row would not feel quite the same, even though the same general structure would remain constant. Interestingly, Dr. Aldebaran stated that consistency was the biggest thing when it comes to fairness in assessment. Thus, she keeps copies of past graded assignments and uses them as a basis for how she grades the next semester. She does this to maintain a level of consistency and fairness and to ensure that similar units would get the same grade regardless of what semester they were written in. So, while the course changes, the grading remains the same.

The second way courses change over time is the result of updates in research. Dr. Polaris said that he looks at his course pack every year and changes a few readings to reflect what is new in

the field. As he tweaks his course each year, some of the readings change and presumably the assignments connected to the readings are updated as well. Similarly, **Dr**. **D**eneb said that she updates her course as a result of what she learns from a yearly conference on technology. At the conference she pays special attention to what technology teachers are currently using in the field, and she works to incorporate these new ways (and presumably remove the old ways) into the new semester. It was from this conference that **D**r. **D**eneb decided to design the Fake News assignment. **D**r. Aldebaran also updates her course based on what is happening in the classroom. As part of her job, she spends time in local schools and when there, looks for new and exciting ways that teaches are teaching. She then works to incorporate these new ways when she has her **PSTs** design their unit plan.

Just because the course is taught by the same person does not, then, guarantee that the course will remain the same. Just as a course changes when it is taught by a different instructor, over time, a course taught by the same instructor changes and adapts as well. These changes come both from responding to the PSTs enrolled and to the ever-changing field of knowledge.

## How does the order of learning affect what is taught and graded?

One thing that really stood out when looking at all my courses was how the order of courses that a PST takes at GU influenced what could be expected, taught, and assessed in each course. I was looking at courses in the middle years of the program, after the general education requirements but before the disciplinary methods courses or student teaching, and this led to some interesting effects.

First of all, the instructors all assumed that the PSTs had already started thinking about their teaching philosophy and knew about the basic tenets of teaching. They believed that the PSTs should already know that teaching is a big job and a great responsibility and were feeling prepared to take on the task.

Second, as the courses were sequential, they also assumed that the PSTs would know the material that came from the course before it. Course A expected that PSTs already knew how to write lesson objectives and had ideas about how to design lessons, and Course BD expected that PSTs knew how to write lessons and could write assessments.

Course C typically came first, but the core assignment in that course required the PSTs to think about how students would be assessed in the unit they designed, so Dr. Aldebaran had to spend some time in her class going over the basics of assessment so that they could complete that component. She could not, however, go into much detail and thus had to grade the students accordingly. Determining content criteria for their assessment was one of the areas where Electra and Maia scored lowest on their PPp1. This makes sense because they really had limited experience learning about assessment. This component was, however, only 2% of the full PP, so while it was assessed, it was not a huge influence on the overall grade.

One place it was apparent that the PSTs had not yet taken a disciplinary methods course was when they struggled with defending the rationale for their PP unit. Both Maia and Electra were given comments on how to fix their rationales (although only Electra lost points), and since they had not yet taken a methods course, they did not have much experience in knowing how to connect their desired unit to their future students. This component was worth 3% of the PP. The PSTs also struggled to connect their unit goal with the standards (worth 3%). It was not immediately apparent if this was something they should have been learning in Course C or would come later. As I did not get to analyze PST submissions of the PPp2, I was not able to find other areas for where PSTs struggled, but from just the PPp1, I was able to determine that it was likely that at least 8% of the full grade came from knowledge that might actually be taught after Course C was completed.

Perhaps because Course A can be a co-requisite with Course C, or because it was not the focus of the course, or because the PSTs do not yet have disciplinary methods knowledge, PSTs were not graded on the realism of their hypothetical course for which their summative assessment was designed. The PSTs needed to create the frame of a unit just to base their assessment on something, but how good the unit was or how well the test was aligned was of little to no importance to the grade. In the rubric, they were only graded on how "clear" their introduction was (A1, p. 3). Course A also clearly came prior to a course on special education, because when the PSTs were asked to make hypothetical accommodations, it was made explicit that these accommodations would rely on "intelligent speculation" (A1, p. 4). Thus, there were areas in the core assignment where I could see that order of learning was taken into account.

The fact that Course BD comes after Courses A and C but before the disciplinary methods courses and student teaching courses was visible not just from knowing the order of courses, but through looking at how the PSTs were assessed. From the rubric I analyzed, as well as from the portfolio and the assignments hyperlinked within, there appeared to be a general understanding that PSTs could and should be able to think about how to design a lesson (from Course C) and that they would include assessment checks (from Course A), but that they were not being assessed on knowing how to teach specific content material. PSTs were to create lessons and demonstrate knowledge of how to use the technology, but the content of these lessons and assessments was not evaluated. One example of this is Fake News submission. PSTs were to design a task using a video, but the content to include was given in the module. Looking at the grading criteria for this assignment, it was graded on being age-appropriate (which would come from the developmental

courses prior to the middle stage), including two activities that would allow students to demonstrate their understanding (which was a mix of Course C and Course A), and using technology was clear and appropriate (a Course BD requirement). Not assessed is the ability to coming up with material to teach the students, which makes sense seeing as this has not been taught yet. This rubric does however demonstrate that knowledge is cumulative in this program, so what comes before is not only a pre-requisite for taking the course, but is also fair game for assessment.

Looking at the SP, it appeared that three major things mattered for success. One, the PSTs needed to properly understand the standards for the course. Two, PSTs needed to be reflective about their learning and be able to demonstrate through their reflections how they have grown in meeting the standards. Three, PSTs need to be able to use academic technology to present material clearly. Thus, from the portfolio, the material learned prior to the course, and to be learned after, was not really assessed. The prior material came in the semester assignments, but for the core assignment, the focus was mostly on what was purely taught in this semester. In this way, the order of learning was less important.

#### Focus on structure.

Looking across all the courses, however, what demonstrated the order of learning most of all was the general focus on how something was done structurally. This structure included both the format in which something was presented and demonstrating that time and effort were exerted. There appeared to be a common theme of caring about what was done and in what manner, more so than what was actually said. There were a few ways this was done, with rule following (to be discussed as its own question later), and with an emphasis on format.

In each core assignment, PSTs were expected to demonstrate that they understood the process for doing something. In Course C, this presented itself as following the procedure for

developing a unit. That all the components were present and that the PSTs demonstrate that they spent time thinking about the unit seemed to be what mattered most. In Course A, this presented itself similarly. PSTs needed to analyze examination data and build a summative assessment, but the focus of the grading was on spending time analyzing and following assessment blueprints, more than if the hypotheses were correct or the assessment assessed the appropriate material. In Course BD, PSTs developed a portfolio to demonstrate that they knew how to think about the standards and promised to keep growing in how they used technology. From the interviews, I learned that the focus was on thinking about which technology to use when and for what purpose. Thus, again, the focus was not on what really PSTs claimed they would teach their future students, but that it would be taught in the proper manner.

Across all the cases, the focus of the core assignments (and the course) appeared to be on being able to go through the steps of behaving like a teacher. As long as the structure was there, there seemed to be an implicit hope that when combined with the disciplinary methods courses later, PSTs would be able to meld what they learned then and now together to become a successful teacher. This hope is likely a result of where the courses are in the program and the expectations for PSTs at this point in their career. In none of the courses were the PSTs ever watched in a classroom as the teacher or asked to model teaching in some other form. While PSTs in Course C and Course A spent a bit of time in the classroom, for C there were field hours and for A it was the AP, the purpose was not the same. As seen with the AP, even this minimal classroom time quickly brought in variance based on the relationship between the PST and the CT. Thus, even if the instructors wanted to know what the PST would be like as teacher and how they would implement a technique or strategy, the opportunity to assess it fairly was not there. Thus, it appeared that all the instructors instead relied on the proxy of developing the structure as the only useful indicator.

## What does a grade in the course tell us?

Most schools give grades, regardless the level. I remember my mother reading me my report grade in my early years, when I earned grades of "E" for excellent and "VG" for very good. I even have a report card from preschool where I was graded on my ability to alternate feet on the staircase and to use scissors. In the teacher education program at GU, PSTs receive alphabetized grades, with A being the top and F being a fail. What these grades actually mean, however, is a bit more complicated.

I took some time looking at the conversion charts from all the syllabi of the courses I studied. I then used these charts to build the one below:

Table 8.1 Conversion Charts								
	C - Aldebaran	A - Polaris	BD - Altair	BD - Deneb				
100	А	Α	Α	Α				
99	Α	Α	А	Α				
98	Α	Α	А	Α				
97	А	А	А	А				
96	А	Α	А	Α				
95	Α	Α	А	Α				
94	A-	А	А	Α				
93	A-	А	A-	A-				
92	A-	A-	A-	A-				
91	A-	A-	A-	A-				
90	A-	A-	A-	A-				
89	B+	B+	$B^+$	B+				
88	B+	B+	$B^+$	B+				
87	B+	B+	$B^+$	B+				
86	B+	В	В	В				
85	В	В	В	В				
84	В	В	В	В				
83	В	В	В	B-				
82	B-	B-	B-	B-				
81	B-	B-	B-	B-				
80	B-	B-	B-	B-				
79	C+	C+	C+	C+				
78	C+	C+	C+	C+				
77	C+	C+	C+	C+				
76	C+	С	С	С				
75	С	С	С	С				
74	С	С	С	С				

Table 8.1 (Continued)								
	C - Aldebaran	A - Polaris	BD - Altair	BD - Deneb				
73	С	С	С	C-				
72	C-	C-	C-	C-				
71	C-	C-	C-	C-				
70	C-	C-	C-	C-				
69	D+	D+	F	D+				
68	D+	D+	F	D+				
67	D+	D+	F	D+				
66	D+	D	F	D				
65	D	D	F	D				
64	D	D	F	D				
63	D	D	F	D-				
62	D-	D-	F	D-				
61	D-	D-	F	D-				
60	D-	D-	F	D-/F				
59	F	F	F	F				

The first thing I noticed was that it was not the same conversion chart for every section I studied. However, there was one area of consistency, which is that As were in the 90s, Bs were in the 80s, and Cs were in the 70s. Thus, no matter how the +/- changed for each grade, the letter was covering the same range.

The second thing I noticed was that while earning the A in Aldebaran's section was the most challenging, her B+, C+, and D+ had larger ranges than all the other sections. Thus, even though getting the A might be most limited, only an 86 in her class would lead to a B+, while all the other sections would call this just a B.

Third, Dr. Altair did not have any way for the PSTs to get a D in her course. Once PSTs were below the C line (which is the grade required for credit), PSTs failed. Interestingly, all the sections I studied did require the PST to get a C in order to pass, but the other instructors still had conversions for a D+, D, and D-. This left me wondering what the difference is between a D and an F, and if perhaps PSTs ever would get in the sixties and not retake the class, but instead switch majors. Answering this question will be saved for another project.

Another way to think about these conversion charts is to consider which letter grades have the largest ranges, which may indicate that these grades are given out most frequently. To consider this, I looked at what percent of above 60% each letter grade represents. I came up with the following chart:

Table 8.2 Grade Bands								
	C - Aldebaran	A - Polaris	BD - Altair	BD - Deneb				
Α	15%	20%	17%	17%				
A-	12%	7%	10%	10%				
<b>B</b> +	10%	7%	7%	7%				
В	7%	10%	10%	7%				
B-	7%	7%	7%	10%				
C+	10%	7%	7%	7%				
С	7%	10%	10%	7%				
C-	7%	7%	7%	10%				
D+	10%	7%	0%	7%				
D	7%	10%	0%	10%				
D-	7%	7%	0%	7%				

From here, I again color coded the cells with green being the highest and red being the lowest. In Dr. Aldebaran's course, 15% of students would get an A, 12% would get an A-, 10% would get B+, C+, and D+ respectively, and then the remaining grades would have 7% of students in each. In Dr. Polaris's course, 20% of students would get an A, 10% would get a B, C, and D respectively, and then the remaining grades would have 7% of students in each. In Dr. Altair's course, 17% would get an A, 10% would get an A-, B, and C respectively, 7% would each get B+, B-, C+, and C-, and the remaining 24% would fail. In Dr. Deneb's course, 17% would get an A, 10% would fail. In Dr. Deneb's course, 17% would get an A, 10% would fail. In Dr. Deneb's course, 17% would get an A, 10% would fail. In Dr. Deneb's course, 17% would get an A, 10% would get an A, b, C-, and D each, and the remaining grades would had 7% of the students in each. Essentially what this tells me is that all the courses are quite varied in their conversions. Grades are not normally evenly distributed, but this calculation helped to highlight the differences in conversion charts.

As helpful as conversion charts can be however, what matters more (at least to me) is how these grades can be achieved. From my understanding, there are two major components for what makes a grade, instructor expectations and deductions or the possibility for revision. In terms of expectations, as discussed in the interview sections of the course chapters, each professor had a slightly different take on what it meant to be an **A** or **C** student in their course. Each, in addition to considering point values, had a qualitative description of what the grade letters meant. These conceptual understandings likely influenced how they determined which letter or number grade to give their students and influenced how they designed their rubrics.

When considering the rubrics themselves, it was interesting to note the differing anchoring methods that each professor used. Dr. Polaris appeared to start with assuming that if PSTs included all the required components, they would receive 100%. Thus, it appeared that he started with full marks and then deducted where necessary. It is interesting, then, that he also has the largest A range. Dr. Aldebaran's rubric, on the contrary, states that what is included on the rubric is what is needed to get a C. This suggests that meeting the rubric will get a C, not meeting it will be lower, and then when PSTs go above and beyond expectations, their grades will rise. Her rubric, then, starts at nearly the minimum, and then adds points when necessary. This difference in anchoring leads to a different understanding of how each grade is assigned and how the instructor thinks about grades. It appears that Dr. Aldebaran believes that meeting expectations means a C, while meeting expectations for Dr. Polaris means an A. This understanding, however, is subjective and not necessarily correct. It is not just the anchoring that matters, but how the rubrics are described. It might be that Dr. Polaris is not writing his expectations in his rubric, but instead describing what is above and beyond. It is also interesting that despite saying her rubric represents a C, the submissions I saw were A and B submissions with points deducted. Thus, sometimes the

idea is not the same is what actually happens. Course BD has a three-tiered rubric with clear levels, so anchoring does not apply in the same way when looking at that particular document. There is still anchoring, however, in other parts of Course BD. For example, Dr. Altair mentioned that if a PST is to "meet all course objectives" for an assignment, then they earn an A. This then anchors the grade at the top.

Another way that grades can mean something different is if there are additional deductions or allowing for revisions. When a course only allows a single submission, there is the expectation that the work submitted at that time is as good as it will be. Almost all my syllabi had some mention of expecting that work be turned in on time and in its best form. However, all the instructors had some mention of revisions or late work, as well. For revisions, sometimes PSTs were encouraged to submit drafts early for review by the instructor (like for Dr. Polaris or Dr. Deneb). This would allow the PST to attempt a draft, get instructor feedback, and resubmit before the deadline, with presumably a much stronger submission than would have been submitted otherwise. In a different way, Dr. Aldebaran allowed PSTs to resubmit core assignments if they scored lower than a C, providing the opportunity for a second chance of success. These allowances to get feedback and make changes alters a bit what it means to get a grade in this course. Thus, a grade is not purely dependent on how well the PST learned the first time, but also on how much they could improve when given individualized help by the instructor. It is possible, however, that these differences in how the instructors allow for revisions account for some of the differences in the grade conversion. Since Dr. Deneb allowed her PSTs to get feedback on the SP before the final submission, it could explain why her cut score for passing is higher than Dr. Altair's.

When points are deducted for something other than the quality of the content, this is also a way that the meaning of the grade changes. For example, with deducting points for late work, the

grade becomes influenced not just by quality of the submission, but with the adherence to keeping a good schedule and being able to turn something in on time. A PST can turn in A-quality work, but if it is turned in late, it will not necessarily get an A. How many points are deducted varied by instructor, which also added some variance into what a grade means. All of the instructors also allotted a portion of their grade to some sort of adherence to guidelines either spelling and grammar, like Dr. Aldebaran, or font and spacing, like Dr. Polaris. This then meant that grades in their courses were not just about how well the PSTs learned the material and could demonstrate it, but also were grading an additional skill.

In all these different ways, what it meant to get an A on an assignment or in the course varied. Whether the actual grade matters or not is a different question, but what I can ascertain is that an A- in one section or course does not always have the same meaning in another. Thus, grades give a general idea, but cannot be translated into a concrete understanding of what the PST knows and can do.

# What the purpose for assessing dispositions, especially professionalism and rule following?

What perhaps stood out the most to me as I looked through my cases and my source data was the strong emphasis on measuring the PSTs dispositions in the form of character and behavior. In particular, there was a strong emphasis professionalism and rule following. When I went through my own chapter analyses, I made 230 notes about what was happening in the chapters and on what was being discussed. Of these 230 notes, 9 of them were about professionalism, 36 were about rule following, and 13 were about in-class behavior. Combined, this means that at least 25% of my analysis was about how the PST should act, rather than what they should know. There are a number of reasons why there might be such a high priority on these behavioral skills. First, teaching is a profession, and as such, there are certain behaviors that will be expected of the **PST** when they enter the field. It makes sense that the instructors in the middle of the program would want to start instilling these behaviors in their classrooms as a way to prepare the **PSTs**. Second, the instructors are teachers and want to be able to run their classes efficiently, so some of these rules or guidelines are about helping the class run smoothly to maximize learning for all. Third, assessing behaviors may be a proxy for the underlying skills that instructors wish to be assessed. In a teacher education program, like in many fields, we cannot always see the skills we want to measure, and thus we need to use something as a proxy. If a **PST** can follow a template for designing a unit plan, for example, this may be a good indicator that they can write a unit plan. Thus, the focus is seemingly on following the guidelines of the template, when the skill that wants to be assessed is unit planning.

The above paragraph lists my hypotheses, built of the data and from prior experience and research. This chapter, however, is about answering questions based on the data, so I will now look at my coding more succinctly to guide you in how I came to these hypotheses. While I coded my data, I acknowledged that the distinctions among professionalism, in-class behavior, and rule following were fuzzy at best, although I tried my best to keep them separate. Therefore, I state here formally that my categories are not fully distinct and mutually exclusive, which is partly why they are all discussed under the same broader question.

To code for professionalism, I looked at expectations about PST behavior throughout the course, although not necessarily about in-class behavior. In this category I found three major components. The first was just about how professionalism would be graded. When the focus was on general grading, I did not include it in this code, although I could have, and would have had an

even higher weight in my count. Courses A and BD were explicit about grading professionalism, so they were counted in this code, but Course C was more implied, and thus not included. The second part of this code was about scheduling. PSTs were advised on how to plan their time, either explicitly, as in Dr. Deneb's syllabus where she told PSTs how much time they should plan to work per week, or more generally, as including a course calendar or recommending that PSTs should keep on top of the course website. The third part was about resources, when the instructors told the PSTs how and where to find the computer lab, and offered suggestions about how to stay organized. Thus, under this code, it was clear that the course instructors not only cared about what material would be turned in or how the PSTs acted in class, but had an interest in helping the PSTs to maintain a degree of professionalism in all that they did with respect to the course.

I coded separately for in-class behavior and expectations. This category included what PSTs should plan on bringing to class (books, electronics) and at the same time, how to not use these items in class as a distraction. There was information about how to work with peers, how to participate, and how to engage respectfully in classroom discussion. Also in this category was how to deal with absences and tardiness. Thus, in this category, the focus was on how to be an active and contributing participant to the course, to enhance their own learning as well as to help and respect their peers. From this I was able to see that the instructors valued how PSTs would behave and interact within their classroom.

The last code was about rule following. In here, I marked all the times I found that the instructor dictated or guided how the PST should submit an item. For example, in Course A, PSTs needed to include a blueprint, a scatterplot, and analyze a number of items. In Courses C and BD, PSTs were given a template to follow for their core assignment submissions. In this last component, I was able to see how strong of an influence following rules influenced the success of a

PST in one of these courses. From the core assignments, I was able to calculate the relative weight of following rules. In the PP for Course C, rule following and mechanics were worth 16% of the entire grade. For the ADP in Course A, 50% of the rubric elements mentioned rule following or mechanics (assuming that all stated components are weighted equally, which I had to assume because I had no information otherwise). For the SP in Course BD, 9% of the grade was from this. While the exact percentage varied greatly, it was still clearly influential. At a minimum, following rules and using correct mechanics was the difference between a letter grade, and at a maximum, the difference between passing and failing.

In addition to coding my own analyses, I also looked at the syllabi descriptions themselves. I wanted to know how much of the syllabus was dedicated to talking about these three components. As a proxy for importance, I calculated percentage of the syllabus words dedicated to describing, explaining, or enforcing components of professionalism, as mentioned earlier in this chapter. All estimates are approximate and dependent on verbosity, but they still provide a good baseline. Course A spent 42% of the syllabus and Course C spent 31%. For Course BD, Dr. Altair's syllabus spent 25% and Dr. Deneb's spent 19%. Additionally, Course C made it explicit the ways that a PST would lose points on assignments for not following the rules.

So, what can be understood from all this? I cannot come to any conclusions, but I can make a few conjectures. First of all, it is possible that the focus on professionalism in the syllabus is less needed the farther into the program one looks. As can be seen from the percentages, there was a decrease in syllabus wording relating to professionalism that corresponded with the order that PSTs take these courses. Perhaps frontloading professionalism in the earlier courses helps to decrease the need for it later on. Another possibility is that as the numbers are all approximate, we can say that on average, a syllabus in the middle years of the program spends about 20-40% of the words dedicated to describing professionalism. This range approximates the range of what I found, and shows that professionalism does take up a considerable amount of space and importance. This is telling because no course explicitly claimed to assess professionalism more than 10% of the overall grade, so this mismatch is noticeable. Thus, despite the initial claims in all of the syllabi about grading, adherence to professionalism is paramount. Perhaps this is because the instructors believe that how PSTs behave influences how they learn, and thus they feel it is important to include.

In conclusion, I repeat the quote from Dr. Polaris, "Some things are indeed hard to assess" (transcript). In some respects, measuring dispositions, at least the internal and mental aspects of it, are quite challenging. What is not hard to assess, however, is how well a PST follows a template or uses proper grammar to express themselves. Thus, I find it unsurprising that when deciding to assess PST progress, the rubrics rely heavily on following directions. If a PST does not follow the directions, it may indicate that they did not understand the assignment or were not willing to put in the effort to grow and change. While it may seem superficial to assess the professionalism, perhaps it is the best proxy we have.

#### Chapter 9: Tensions, Philosophy, and Implications

As mentioned in my Chapter 1, my interests lie broadly in assessment and fairness. I conducted this study because I felt that teacher education was a context in which to study how assessment was used, but I also felt that it would be a good springboard for a broader and deeper conversation. Assessing teachers, especially in the middle of their program, is challenging and important, and as such, it provides an appropriate setting for my philosophical questions.

As a result of the data that I collected, I was not able to study in depth many of the tensions that I wanted to see, although they did show up in places. To really understand a range of assessment tensions, one would need to collect a myriad of additional data, including PST interviews and data from a larger sample size. In this chapter, I describe and explain the tensions that were present in my findings, and suggest ideas for future research to systematically examine these tensions further. I present these tensions as sections, but as in the prior chapter, there exists some overlap and the sections are not mutually exclusive. These tensions all relate to issues of fairness, but from slightly different perspectives.

Before I discuss the tensions explicitly, I begin with an overview of how I see fairness in assessment. I then transition to discussing the tensions I found in my data. The first tension is related to how the core assignment is related to the curriculum, and how this relationship influences fairness. I start here because as the core assignment influences the curriculum, many of the other tensions stem from this point. Second, I look at tension that develops from the order of the courses at GU. Because these middle courses are not subject specific, I explore how subjectarea knowledge may be related to what is learned or taught in the courses. Third, I discuss how dispositions are assessed in the three courses. In addition to assessing teacher knowledge, I found that there was a strong emphasis on assessing dispositions. Fourth, I examine what, other than

teacher knowledge and dispositions, is assessed in the three courses. From my data, I found that English fluency and access to resources were assessed, as well. Lastly, I describe the number of tensions that I found surrounding the rubric and its development.

Before concluding the chapter, I revisit my three research questions and concisely answer them based on what I found from my data. I began my project wanting to know:

- 1. What are PSTs expected to learn, do, and know in the middle years of their teacher preparation programs?
- 2. How do teacher educators assess these middle-program PSTs?
- What are the tensions involved in these expectations and assessment decisions?
  I use this final space to share what I learned about my questions.

#### Fairness in Assessment

There are two common tropes in life. The first one is "life is not fair" and this is often used as a reason to explain inequalities. The other one is that "life should be fair," and this trope is used by those who believe that things can be different. What it means to be fair, however, and how to make things fair, changes based on perspective and experience.

To start more broadly, and outside of education, there is a national debate about the Electoral College and its place in the American voting system. Proponents for the Electoral College argue that this body helps maintain the balance of power between the larger and smaller states, and ensures that the voices of those who live in less populated states still get heard in the election. If voting were based on pure numbers, those who live in large cities would outnumber those in rural areas, and would be able to vote for politicians or policies that support urban rights at the expense of rural rights. Opponents of the Electoral College argue that its existence is unfair to people who live in populated areas because their individual votes are worth less. Because of the weighting system in the Electoral College, a vote by a citizen in Montana is worth three times the vote of a citizen in California. Proponents for the Electoral College argue that this is then unfair to people who live in more populated areas, because the vote of the individual is worth less and therefore this leads to their needs and voices being muffled. I find this debate interesting because both sides are fighting for fairness. Both sides want their voices to get an equal space in the election and worry about others getting an undue advantage. Thus, people often are fighting for fairness, just with different conceptions of fairness.

Fights for fairness also exist in education, and including in teacher education. People want things to be fair, but how, and for whom, is the question. When preparing teachers, what really is the goal and how do we know it is being met? Is "teacher knowledge" what is important? If so, how can we construct a test of teacher knowledge that does not privilege certain knowledge, and thereby be fair from CIV? Every test format leads to decisions that must be made, and each decision may influence who succeeds. For example, Dr. Aldebaran has strict rules in her course for use of grammar and spelling, but what counts as proper grammar and spelling is dependent on background. Is deducting points for not using Standard English disadvantaging those who come from different backgrounds, or is knowing Standard English a necessary skill for being a successful teacher in America? Depending on who you ask, the answer will vary. Different answers will lead to different ideas of how fair Dr. Aldebaran's grading system really is.

The debates go even further. There are many types of teacher knowledge, as described in the literature review chapter. Specialized Content Knowledge is different from Knowledge of Content and Curriculum. Deciding how to measure each and how to weight them is a choice that must be made in a teacher education program. The results of this decision, however, may influence who succeeds. What really is the right mix of the different types of knowledge and how

do we know that this mix is right? Does fair mean the same standard for everyone, or does it mean that every **PST** has an opportunity to succeed in their own way?

As I proceed through this chapter, I describe many of the tensions that presented themselves as I conducted my research. I do not have answers or solutions to the tensions, but I present the ideas and suggest methods for gaining further insight and understanding.

# How are the core assignment and the curriculum related? How does this relationship influence fairness?

In each of the three courses I studied at GU, there was a core assignment. To pass the course, PSTs needed to not only get a C in the course, but also get a C on the core assignment. The core assignment was an agreed-upon assignment by the department, although I did not get any details on how it was created. Nevertheless, I learned that this dual passing requirement was intended to ensure consistency between sections. The goal was that the core assignment would provide a common measure for all PSTs taking the course, regardless of instructor or term. Thus, the PST would be assessed in the same way in every section and in every year.

I was also able to get a bit of information on how this core assignment related to the course. In this section, I look first at when the core assignment influences curriculum and how this affects fairness. Then, I look at how the curriculum influences the core assignment, and how that affects fairness.

### When the core assignment influences the curriculum.

Dr. Aldebaran said that she designed her course using the theory presented by Understanding by Design (transcript). The purpose of Understanding by Design is to focus on the required outcomes of the course, and then to plan the course accordingly. Thus, as the core assignment is the one of the major assessments of the course, Dr. Aldebaran used it to design a

curriculum that focuses on helping the PSTs learn the required material to succeed on the PP. In this way, the core assignment influences her curriculum decisions and modifies how and what she teaches in order to align to the core assignment.

Additionally, having a core assignment can influence what material gets learned in the course and can lead to other important content being permanently cut. When the core assignment dictates the curriculum, it may narrow what is taught, as the instructor works to ensure PST success. This phenomenon has been noted in K-12 education, as well (e.g. Au, 2007 & Berliner, 2011). If this narrowing means that the instructor will only focus on skills that are essential to being a teacher, this may not be problematic. It may lead to fewer tangential topics and more emphasis on critical knowledge and skills for good teaching. However, narrowing does not necessarily mean that the course is focused on the right material. The core assignment could influence the curriculum to focus only on a narrow portion of the knowledge and practices that a PST needs to learn (e.g. that a PST should only ever teach using Jigsaw lessons), and hinder the PST's ability to become a well-rounded teacher.

Nevertheless, when the core assignment influences how the curriculum is designed across sections, then one can expect that PSTs taking any section of the course are learning the same material. This then creates a degree of consistency and helps the program understand what is taught and assessed in every course.

### How does the core assignment influencing the curriculum affect fairness?

With this directionality of the core assignment affecting decisions about curriculum, I was curious about how this core assignment could also influence fairness. On the one hand, having a common assignment across all sections does provide a level of consistency. It allows the program to make claims about the **PST**s and to be able to show accreditors that they know what is being assessed in each course and section. By having the same assignment across sections and over time, all PSTs are held to the same standard. The same standard can mean fairness.

On the other hand, just because there is consistency, it does not necessarily mean fairness. If the core assignment is biased in any way, it remains biased in all sections and over time, and can contribute to the course curriculum being biased. This bias could lead to certain PSTs leaving the program due to "poor performance" which actually means a failure to comply with biased standards. If the core assignment is discriminating based on a necessary teacher skill, then this discrimination (in the mathematical sense) is appropriate. If this discrimination is based on something else, however, like specific content knowledge that not all PSTs have learned yet, or on personal characteristics that are separate from teaching (like how artistic a PST is), then this can be problematic. Thus, having a core assignment is only fair if the core assignment is overall fair and constructed to measure knowledge and skills that are important to becoming a teacher.

Additionally, fairness comes into play when the core assignment influences how an instructor can respond to individual PST differences in the course. If the core assignment creates a narrow and rigid curriculum, then instructors may feel compelled to follow the curriculum regardless of who is in the course. This would then eliminate the options like the pre-test administered by Dr. Altair to adapt the curriculum to the needs and skills of the current PSTs enrolled in the course. Furthermore, this then means that the core assignment, which acts as an assessment, is not aligned to the needs and growth of the given students. Depending on how it is written and scored, it is vulnerable to issues like the ceiling or flooring effect, which caps PST growth or truncates misunderstanding and does not show true levels of understanding (Baker et al., 2010). When an assessment is not aligned with the PST understanding, the results that can be

gleaned from the grades are limited. Thus, both the assessment results lose meaning and the instructor cannot cater to **PST** needs.

Additionally, the core assignment influencing the curriculum can be an issue of fairness if, for example, the core assignment requires PSTs to only use literature from a prescribed list. In this example, the course instructor would likely also teach mainly from this list, and other important sources and authors could be ignored. Other equally important articles and readings would be then eliminated from the curriculum and never taught.

For future research, I recommend looking more deeply into the core assignments and determining how fair they really are. Ways to consider this would be to look at several years of PST scores and run DIF calculations to determine if certain non-important teacher characteristics are being screened. Maybe the assignment is especially challenging for PSTs who are from rural areas, or who are focusing on secondary science education. Getting more numerical data can help clarify this degree of fairness. Qualitatively, researchers might want to conduct focus group interviews to better understand how different groups of PSTs understand and interact with the core assignment. Additionally, it would be useful to conduct research on the predictive validity of the core assignment. How predictive is success on the assignment of future teaching ability?

#### When the curriculum influences the core assignment.

On the other hand, Dr. Altair said that she used her academic freedom to add components to the Course BD core assignment, modifying the SP to better align to the focus she gave the course that semester. Dr. Altair would adjust the directions of the core assignment to match how she focused the course that semester. Then, as she made adjustments to the instructions, she made sure to make corresponding adjustments to the rubric to reflect these changes. From this direction, it was the course that influenced how the core assignment was graded. Thus, while at times the core assignment might influence what is taught, in this instance, what is taught might influence how the **SP** is assigned and graded.

#### How does the curriculum influencing the core assignment affect fairness?

One may believe that modifying the core assignment toward what was taught and focused upon in a course would make the core assignment more fair for the PSTs in that particular section. I agree that making a critical assignment aligned to the taught material is important. The tension appears when considering what happens across the course when one section makes such a change. As the core assignment changes from section to section, the claims that can be made about the PSTs passing the course change, as well. As the assignment and rubric are no longer the same, comparisons cannot as easily be made between sections. What it means to pass the course also changes. This becomes especially important if two sections make similar changes to the curriculum, but only one section alters the core assignment to reflect these changes. Now, results from the core assignment rubric are even more complicated. Presumably, the PSTs who took the altered curriculum with the altered core assignment will perform better than those who do not get the altered core assignment, because the alignment should make the assignment easier. Now, the claims that can be made from the grades are even more skewed. While adjusting the core assignment to match the focus of the course leads to fairer assessment for the PSTs in that section, it leads to issues elsewhere.

To research how this alteration affects fairness, I recommend a study comparing the PSTs' scores between sections that alter the core assignment and those that do not. I would recommend a detailed task description analysis by experts, such as researchers who study knowledge in teacher education, as well as psychometricians, to understand what is being taught and assessed in each

section and how this changes the difficulty and fairness of the assignment. I also recommend, if possible, randomizing who gets what core assignment and seeing how this influences performance.

### Considering both directions.

In both situations, when the core assignment influences the curriculum and when the curriculum influences the core assignment, issues of fairness surface. Questions arise about what is the fairest way to assess PSTs and how to ensure that claims made about a course are reasonable and important. This becomes especially critical if it leads to a narrowed curriculum or a biased section of a course. Interestingly, Au (2007) found that "as teachers negotiate high-stakes testing educational environments, the tests have the predominant effect of narrowing curricular content to those subjects included in the tests" (p. 264). Berliner (2013) found similar results. At GU, the "high-stakes testing" was the core assignment, and while in some instances narrowed the curriculum, like for Dr. Aldebaran, it did not narrow it in Course BD. Thus, it appeared that either the core was not as high-stakes as can be found in K-12 environments, or that there is something about academic freedom in higher education that prevents this phenomenon from being the norm. My project only touches on this issue of the effects of a standardized assessment and its influence on the curriculum and fairness, but it brings up an important springboard for future research.

#### Does subject area matter in non-disciplinary courses?

All three of the courses I studied were non-disciplinary courses and were taken by all PSTs at GU. There were two tensions that presented themselves as a result of the setup. The first tension was discussed a bit in Chapter 8 when I discussed the focus on structure over substance. I will say more about that tension in this section. The other tension relates to how different subject areas might influence success in a course or on a specific assessment.

#### Structure versus substance.

To begin, I discuss in more detail the tension of structure versus substance. I define structure as the building blocks of teaching, such as how to write course objectives or format a summative assessment. Structure in this case is somewhat removed from content knowledge or age-level, although it cannot be completely devoid or else it loses meaning. On the other hand, I define substance as what goes into the structure. Here, substance would be something like what the learning goal actually is or the content being measured by the summative assessment. Because the courses I studied all came before the disciplinary methods courses, there was a focus on assessing structure, and necessarily, PSTs were seldom graded on the actual substance.

This structure versus substance situation is interesting because it assumes a number of things. It first assumes that one can understand the structure without fully knowing the substance. For example, it assumes that a **PST** can truly understand what it means to write a course objective without thinking about what the course is or what it might mean in practice for a lesson to be centered around a particular objective. Thus, when the structure is assessed, it is assessed essentially in a void. The second assumption is that learning the structure without the substance will lead to the ability to combine the two once the substance is learned. This assumption relies on the hope that the substance can be inserted later. It is not all hope, as **Dr**. Aldebaran talked briefly about how in the methods courses the instructors teach the **PSTs** how to insert the knowledge, but how explicit this connection is might depend on the instructor. Thus, essentially **PSTs** are assessed on knowing enough structure to indicate that, in the future, they might be able to meld it with substance, while not actually assessing the second part.

This tension is particularly interesting when viewed through the lens of Ball and colleagues' (2008) framework with respect to the different types of teacher knowledge. Knowledge of structure

is embedded in the component of Knowledge of Content and Teaching (KCT). As I explained in Chapter 3, KCT refers to what teachers need know about instructional strategies and resources in order to maximize their ability to help their students learn (p. 401). However, unlike how the skills were taught at GU, according to Ball and colleagues, knowledge of structure and substance works together. Despite this scholarly claim, in practice, it appears that KCT, at least in the middle of the program, is really just KT, with content coming in later.

The pattern of teaching structure courses first and then methods is not unique to GU. A colleague of mine explained his teacher education program to me similarly. First, the PSTs learned content knowledge, along with human psychology and child development. Next, they took pure structure courses. Then, they took methods courses, and finally, participated in student teaching. Because there is only so much time in a course and a program has to be ordered in some way, it appears that this pattern is common. However, I wonder about how common this pattern is and if there are other or better ways to sequence the classes. Furthermore, I would be interested in speaking with Ball and colleagues to hear their opinions on KCT being divided into two separate components. For future research, I recommend reviewing the sequence of courses in different programs and looking at the outcomes as a result of this sequencing. I also recommend looking for programs that integrate this knowledge all together and seeing how that might make a difference. As of where I am now, I do not have a hypothesis on how this might play out, but as it kept showing up in my research as an issue, I think further study could be quite interesting.

### Does the subject-area change the difficulty level?

Also under this question of course order was the issue of how future teachers in different subject areas might differentially succeed or experience difficulties in a course, and consequently be assessed on the wrong construct. Supposedly, all three of the courses I studied were

independent of subject-area knowledge and therefore should have been just as challenging and meaningful for all PSTs. However, according to the instructors I interviewed, PSTs did not always see the courses this way. For example, when talking to the Course BD instructors, it came up that not all the PSTs had as easy a time understanding how they might use technology in their future classrooms. PSTs who focused on science tended to easily find areas where they might want to incorporate technology, but secondary English teachers had to spend more time thinking about how to use it. The instructors believed that all PSTs would be able to effectively use technology to enhance their future teaching, but the immediate link was not there for all subject areas. Therefore, depending on the subject that the PST planned to teach, this course required more or less creativity. It was not that the course itself was inherently harder or easier for certain subject areas, but that previous experience with how the material was taught made it more or less challenging to envision. Thus, while the course was intended to be fair and accessible to all PSTs, there was a bit of a learning curve for some PSTs. The assessments then assessed some of the PST on not only their ability to create technology-infused lessons, but also on their ability to stretch their imagination to think of ways to use technology, while other PSTs were only assessed on their ability to use technology in conventional ways.

I also wondered about the role of content when I considered the AP in Course A. This project had PSTs administer a test in a current classroom and then review the results. I only had the opportunity to look at submissions from two PSTs, and both administered multiple-choice assessments to secondary students. From these two submissions, the difficultly level was fairly even, although even here there were differences. Anwar only had to grade a 20-question test while Yildin had to grade a test that had over 80 questions. If nothing else, test length changed how long it took each to complete this project. Next, content knowledge came into play in this project because the PSTs needed to analyze the results and hypothesize why certain questions were answered wrong more frequently. While the actual hypotheses were not graded for accuracy, knowledge of the content was important for even beginning to tackle the task. One might find that the better one knew the subject material being tested, the better they could do on this project. I wondered if this project would be harder or easier for a PST who chose an elementary assessment to analyze. Would having simpler content knowledge make it easier to analyze, or would it not? And, if it is easier, are the PSTs really all being assessed on the same skills?

I also wondered about the content-knowledge focus of the course instructor and how this could potentially influence grading. Dr. Aldebaran mentioned that she tries hard to not penalize PSTs for presenting a math lesson in a way that she would not. As a math person herself, she found that without checks, she could become more invested in the math PSTs' submissions, and as such, worked to avoid this bias. She also found that it was challenging to grade some of the science PPs because they used content knowledge that she did not have. In order to grade these assignments, she spent time researching the content and consulting with colleagues. While Dr. Aldebaran put in these checks to ensure consistency, did this really eliminate bias? I found that there was a time when both Maia and Electra had similar comments on their submissions, but Maia scored higher. Was this because Dr. Aldebaran was attempting to not be overly harsh on her math students and accidentally graded Maia more leniently instead? This was only one instance, but it brings up an important point about instructor knowledge influencing grading.

I recommend future research look into how content-knowledge background of the instructor influences grading decisions within a non-content-focused course. I would be interested in seeing how the match or mismatch between the content focus of the instructor and the PST plays into how the PST is assessed.

### How are dispositions factored into grading?

Another tension I noticed as I analyzed my data was the prevalence of dispositions, particularly personal characteristics, and their implicit or explicit effects on grading. Explicitly, dispositions were graded when it came to how professionalism was calculated. Implicitly, it appeared in how the professors understood what it meant to be a successful student in the course.

Explicit grading of personal characteristics was visible in the grading of professionalism. In Course C, PSTs were graded on how well they interacted with their peers during class discussion and group work, and how they responded to being tardy to class. In Course A, PSTs were told, "Attendance does not guarantee full credit, active engagement in the class does" (syllabus). However, what it means to be actively engaged was not defined. In Course BD, one of the course objectives was to be a "collaborator," and this was seen in Dr. Deneb's assignment of the group project that is coordinated virtually. In each of the courses, although not in all the same way, PSTs were graded on their behavior as a student in the course. I wonder, however, how well being a good student correlates with being a good teacher. Is an outgoing student, someone who is ready with a hand raised and an eager contributor, also the person who will best monitor their own classroom discussions as a teacher? Is the person who is friendly and gets along easily with peers someone who will understand all student needs? I am not suggesting that these people will not be good teachers, but I am presenting the possibility that a shy and reserved student may also shine in the classroom as a teacher. Someone who is uncomfortable talking with peers and instead is the quiet listener might connect well with students and be able to keep student discussion flowing. For some, teaching is akin to acting, and student persona and teacher persona do not always align. Thus, I am curious if the dispositions exhibited as a student directly translate to teacher dispositions.

Nevertheless, the PSTs are graded on their student behavior in a teacher preparation program. I assume that they are graded in this way because at the time that they are taking courses, they are being students. The instructors of Courses C, A, and BD never get to see the PSTs in the classroom and there is no performance assessment. Thus, they grade what they can see. What they can see is in-class behavior, as well as the ability to follow rules and use specific mechanics in assignment submissions. These three components make up how dispositions are often assessed in the teacher education. My concern, however, is that these elements may not correlate with good teacher performance in the classroom. As there is still a lack of empirical evidence for which dispositions are most correlated with good teaching, there is the chance that PSTs are assessed on the wrong behaviors (Borko, Liston, & Whitcomb, 2007, p. 362).

For future research, I recommend looking into this correlation between dispositions and good teaching and considering how to best weight student characteristics (and their ability to follow rules) when assessing future teachers. Until we have concrete evidence for which dispositions best lead to good teaching, we should be careful about how we assess them.

On the other hand, teaching is a profession and, as a profession, its members are held to a certain standard. Being someone who completes work on time, gives advance notice about absences, and communicates clearly are all skills that teachers use on regular basis. Assessing that a PST can act professionally is not outside the realm of what we would expect from a teacher preparation program. The question remains, though, of how and when it should be assessed. What claims is each course trying to make, and how does measuring professionalism fit in?

To research this, it might be that the best method will be to consider early career teachers. I recommend finding early career teachers who are deemed successful teachers, by the school, by parents, and by their students, and then working backward to understand what kind of **PST** they

were. These teachers can be interviewed or surveyed about their PST experiences and in-class behaviors in their teacher education program. Then the researchers can reach out to these teachers' teacher education instructors and get their evaluation of their in-class behavior, professionalism, and rule following. Using the surveys and instructor data, the researchers can look for patterns and see if these behaviors are in fact correlated, and how strongly. The results from this study can strengthen our understanding about how much weight we should attribute to these "non-teaching" behaviors.

#### What factors, other than teacher knowledge or dispositions, get assessed?

In this section, I look at two non-teacher knowledge or disposition components that seemed to be assessed in the teacher education program at GU. As I quoted in Chapter 1, "current measures for evaluating teachers are not often linked to their capacity to teach" (Darling-Hammond, 2010, p. 2). In this section, I submit two areas where this could be the case. The first is fluency with the English language, especially Standard American English (SAE). The second is access to resources.

# When and how should teacher education programs assess English language skills?

In all three courses, a portion of the grade for the core assignment was allocated to using Standard American English (SAE) vocabulary and grammar. From the data I collected, I did not learn much about the demographics of the PSTs who attend GU. One of the four professors did mention that her course tends to be taken by predominately white females, which would indicate that the program is predominately white and female, and possibly all native English speakers, but this does not necessarily indicate that most of the PSTs grew up in homes that spoke SAE or would be teachers in classroom with SAE speakers. It is possible, however, that this requirement in the rubrics to use SAE may influence who succeeds in the teacher education program at GU.

Does knowledge of SAE make a teacher better, however? Does grading SAE privilege PSTs who come from homes that speak SAE regularly? As mentioned in Chapter 3, the military tests that relied on knowledge of English found that it was misunderstanding the intelligence of soldiers who were non-native English speakers (Boake, 2002). It is possible that this continued reliance on SAE might also be similarly influencing how we assess PST. As we look to expand the teacher workforce, and to prepare teachers to be able to empower all students to learn, what might we be excluded by focusing only on one type of English? I recommend future studies to look further into language usage and how this can help or harm students' learning in school. There is already a field of knowledge dedicated to code switching and validating student home language in the classroom (e.g. Amorim, 2017; Craig, 2016; Lin, 2008; Wheeler, 2004). Perhaps building on that field, and looking into how we prepare PSTs to speak and write in the classroom, can lead to many interesting future studies. I would be particularly interested in reading studies that demonstrate how expanding what it means to speak English correctly could benefit future populations of students. If it is found, however, that speaking SAE does, in fact, make a teacher better, I recommend researching to what degree it makes a difference. There is only so much time to prepare PSTs in a teacher education program, and as such, it is important to focus on preparing PSTs with the skills that will most enhance their teaching. It would likely be problematic if teacher education programs focused on teaching SAE at the exclusion of other skills, such as lesson planning and assessment development.

While I was not able to come to any conclusions or deep analysis on language usage in the PST classroom, it did come up enough to point out that it may be an issue moving forward. As

grading language takes some weight away from content, it is possible that it can bring in CIV. When we assess PSTs, we want to know that we are assessing critical skills and abilities, and that our assessments are not skewed by unnecessary components. Therefore, I suggest more research into the benefits of SAE and how much it should be weighted in a teacher preparation program.

#### What happens when we require access to resources?

In all three courses that I analyzed, there appeared to be a general goal of keeping costs low. Many required readings were available online through the course websites, and other books were kept to a minimum. For the technology course, there were a few additional costs (like for microphones or jump drive), but in many cases there were options for free workarounds, such as Open Office instead of Microsoft Office or the use of the on-campus computer lab. As such, it appeared that in most instances, care was taken to ensure that access to resources would not be a problem.

However, despite the care to make the course accessible to all, there were still a few barriers. According to Luo and colleagues (2011), people enroll in online courses for primarily one of three reasons: level of control, independence, and/or satisfaction (p. 283). They define level of control as the freedom the person has to manage their time and pace in the course, independence as the feeling one gets from taking an online course, and satisfaction as having the learning style match with personal preferences. For the purpose of considering accessibility, I focus mainly on level of control.

If, as the scholars claims, a PST will enroll in an online section of a course because it allows them more control over the timing and pacing of their lesson, then needing to come to campus to use the computer lab, during the hours that the computer lab is open, may be antithetical to the purpose of taking the online section. It could be possible that a PST chooses the online option of
Course BD for the freedom to work at a time that is convenient for them (such as around a work schedule) but then will find that this still is impossible if they need to be able to get to the computer lab. The PST may then struggle to succeed in the course, not because the material is too challenging, but because they cannot get to the necessary resources in order to succeed. If the reason one needs the flexibility is because they need a full-time job to manage the college tuition, then this lack of flexibility also becomes an economic hardship and may discriminate against financially struggling PSTs.

With my research, I did not look for evidence of this being a problem, but it still appeared as a question as I reviewed my collected data. As discussed in Chapter 3, access to resources can lead to CIV, as what is being assessed conflates PST knowledge with what they had the time and money to access. Moving forward, I recommend research into how access to resources, particular computer lab programs, may influence PST success in teacher education programs. As teacher educators, we should be aware if PSTs do not know the course material because of a lack of understanding or because of a lack of resources. We want the scores from the assessments to reflect the former, and not the later. If teacher education is assessing PSTs on their access to resources, instead of on their teaching ability, then this is an issue of fairness. Interviews with PSTs about their experiences, as well as a DIF analysis on scores based on economic background, may be good places to start to researching this issue.

#### What is in a rubric and how does that shape the assessment?

Something I noticed in my analysis was that each course used a very different type of rubric for the core assignment. As a reminder, Course C used a rubric that gave point values per element, but not a point breakdown within the elements. Course C also listed the rubric elements that minimally were needed to pass the assignment. Course A used a rubric that listed what was needed

to get 100% on the task. There was weighting between assignment sections, but no weights given within each section. Course BD used a three-tiered rubric that matched the elements to the course objectives, but was not as explicit about how the individual components of the task would be assessed.

There were four defining features across the rubrics: the anchor, the weighting, the descriptions, and the focus. The anchor is the target behavior, the weighting is the explanation of how points would be broken down, the descriptions are the words used to indicate what needs to be included for the credit, and the focus is the choice for what exactly is graded. Each rubric took a different approach to each of these features, and consequently presented different ideas for what it means to be assessed in the course. I describe each of the four features individually over the next few pages.

### Anchoring.

What one sets as the target of an assignment can change how the assignment is approached by PSTs. Dr. Aldebaran chose to describe the minimum needed for passing and Dr. Polaris decided to describe the maximum. Both choices are valid, and yet they lead to a very different way of using the rubric as a guideline when completing the assignment. (Course BD avoided this debate by including three tiers in its rubric, thereby giving descriptions for both moderate and excellent submissions.) For future research, I recommend interviewing instructors more closely about their choice of target and their reasoning about this choice. Why did Dr. Aldebaran want PSTs to know the minimum requirement when Dr. Polaris wanted PSTs to know the maximum? How did this change how the assignment was assessed? Did one method of grading lead to stronger submissions? What does it mean when a program is not consistent in anchoring? Are PSTs ever confused about which way they should be reading the rubric? Does this influence fairness?

Since all three courses used different rubrics, the assumption was that each serves an important purpose. Dr. Polaris said that he was intentional with his rubric and called himself a "rubric minimalist" (transcript). He explained that making a full rubric in the style of Course BD would in fact limit creativity and freedom, both for himself and his PSTs. By stating only the top objectives, he allowed his PSTs the space to be creative with their submissions, for instead of trying to fit themselves into a particular box, they could consider the top goal and then design a submission that allowed them to express their understanding in a unique and meaningful way. This rubric style, of grading holistically from a top anchor, also allowed him the freedom to grade the submissions based on overall quality, instead of being boxed into a rigid rubric. If a PST really excelled in one area of the rubric, or really did terribly in one part, he could grade this with a relativity that would not be afforded to him in a more stringent rubric. It also allowed him to grade new and different submissions for what they were, rather than how well they fit into prescribed cells. Additionally, although less explicitly stated, this style of rubric suggested that excellence was the goal. As the rubric only listed the top requirements, it suggested that doing everything was what was expected.

Dr. Aldebaran defended her method for rubric design because she felt that the purpose of the assignment was to know if the PST could design a curriculum unit sufficiently to continue on in the program. For her, the focus appeared to be on the concept of "enough." Thus, she designed her rubric with "enough" as the critical area upon which to elaborate. Since she wanted to know if a PST should be able to pass or needed to retake the course, the pass line was be clearly marked and defined. Her approach suggested that the goal of the assignment was to meet competency

requirements, not to determine excellence. Furthermore, Dr. Aldebaran had the teaching approach that if one can do something a few times, that should be sufficient for demonstrating knowledge. She believed that if a PST could write three or four lessons in detail, then that demonstrated that they could write lesson plans. She said she had colleagues who graded up to ten lesson plans, and she felt that was just excessive. Thus, Dr. Aldebaran set up her grading structure with the pass line in mind, and figured that that was what really mattered (even though in practice she appeared to be grading more with the top score in mind).

So what do we make of this tension? Is one way of writing a rubric better than the other? Likely not, although more research might be needed to defend this empirically. What I take from this tension is that rubric design stems from beliefs about the purpose of assessment, and from how one wants a PST to use the rubric for guidance. If the purpose is to determine pass or fail, then Dr. Aldebaran's rubric is the one to use. If the purpose is to place PSTs into boxes along a predetermined continuum, then Course BD's rubric is the one to use. And, if the purpose is to allow PSTs the freedom to express their understanding in a variety of meaningful ways, then Dr. Polaris's rubric is the one to use. Rubric choice, especially with regards to anchoring, should match the purpose. There likely is no best overall rubric design, but there are best ways given the particular purpose and goal for creating the rubric.

## When should rubrics have weighting?

Each of the rubrics used a different weighting structure for point allocation. Course A gave point values for each element in the rubric, but did not explain how those points would be broken down. Course C gave weighting for sections of the core assignment, but did not allocate weight for the given elements within each section. Course **BD** used a tiered system that described the different behaviors necessary to make each point.

While clear weighting is often desired, especially because it makes it apparent what matters most in a submission, it is not necessarily the best method from a teaching standpoint. As I wondered when I have taught courses, is giving too many directions turning the assignment into a checklist? The more I tell my students about how I will grade an assignment, the fewer surprises they should have when I return the graded assignments, and also the better they will understand what I want from them. However, at what point does the rubric end up doing all the thinking for the students? When I give detailed directions, I end up grading my students on how well they follow a list of instructions. A consequence is that I deduct points for not reading the directions carefully and missing a piece, instead of deducting points for not knowing how to do something. There seems to be a delicate balance between not giving enough information and being so clear the assignment becomes similar to a color-by-number.

I found that looking at weighting highlighted this tension in balancing how much direction to provide. In teacher education, we are preparing future teachers who will need to be active thinkers and reflective problem solvers. What is too much oversight and what is too much shadow? My dissertation did not provide any answers on the best way to consider writing rubrics or providing the adequate amount of weighting, but it brought up the question. As each course entertained weighting differently, they presented different theories on what is appropriate. I recommend that future research looks deeper into the balance between providing weights or point allocation to avoid subjectivity and how being too explicit can eliminate the need for **PST** thinking.

## Descriptions: Rule following or subjectivity.

Similarly, how rubric elements are described leads to tension of how much is too much and how little is too little. In the rubrics I analyzed, there was a mix of rule following and subjective descriptions for what needed to be included. Objective descriptors are items like, "all handouts are

attached" (C1, p. 4) or "assessment item choices are aligned<sup>8</sup> with outcomes" (A1, p. 3). Subjective descriptions, on the other hand, are items like "exhibits evidence of ability to design, develop, and evaluate authentic learning experiences and assessments" (B10, p. 2)), or "the introduction is clear and helpful" (A1, p. 3).

When rubric elements are more objective, like those in binary inclusion/exclusion form, they are easier to assess and lead to less ambiguity. Any instructor would be able to use a rubric of this form and grade the assignment even if they had not been the acting instructor in the course. However, these more objective rubric elements say little about the actual thinking ability of the **PSTs** and likely suggest the **PSTs**' ability to follow directions, rather than to be a future teacher.

On the other hand, when the rubric elements are more subjective and are less tied to direction following, it is not as apparent how they will be assessed. What makes an introduction "clear" or "helpful"? Helpful for whom? What needs to be included for it to be clear? Do we mean format, enough information, the right information? What evidence is good evidence for demonstrating an ability? These rubric elements can be hard to describe, and thus can lead to confusion about the expected outcome. Sometimes these items are hard to articulate, but the instructor knows it when they see it, and sometimes it is described vaguely because there are many different ways to do it, and the instructor does not want to limit or overly direct the PSTs.

Just as with weighting, the word choice within the rubric is subject to a delicate balance between being too prescriptive or too vague. I recommend future research delves into sense making around rubric language and studies how instructors and PSTs work together to construct meaning from a rubric. How do both parties know what to expect in a submission with respect to quality? This tension presented itself in my research, and thus I am curious to know more.

<sup>&</sup>lt;sup>8</sup> Alignment can be objective when it is about matching the words and topics. It can also be subjective when one considers alignment to a particular meaning.

#### Focus on standards or components?

Lastly, I found a tension in what should be graded in a rubric. Courses C and A seemed to focus more on grading the submitted components as described, such as looking at inclusion of a required element and then grading it for quality. Conversely, Course BD graded the core assignment from the lens of meeting the course objectives. Both methods are valid, but present a different understanding about the purpose of the core assignment and what the grade means. It appeared that in Courses C and A, there was an implied connection between meeting the core assignment requirements and then being able to meet course standards. While there was no clear link between the rubrics and the course objectives, as passing the core assignment was critical to passing the class, there was the assumption that they were measuring the same ideas. In Course BD, however, it was the instructor's job to make the leap from what was presented in the SP and then to grade based on the standards. Thus, the tension here was on where the leap was and who made the jump. Was it the instructor's job to see the standards over the course of the core assignment and to grade it more holistically through this lens, or was it the job of the department to create a core assignment that would lend itself to this leap when it came time for accreditation? I do not have too much to say about this tension, but that it may be worth exploring further. What changes when who makes this leap changes? What does it mean for the course or for the PSTs' understanding of what they have learned?

## **Synthesis**

Before I conclude, I will take a few paragraphs to concisely explain what I learned with respect to my research questions.

# Research question 1: What are PSTs expected to learn, do, and know in the middle years of their teacher preparation programs?

Based on the three courses I researched at GU, I have some ideas of what is expected of PSTs in the middle of the program. While this is only one university and only three courses, this list is not exhaustive nor necessarily generalizable, but it still provides a general framework. First, there is structural content knowledge expected of PSTs. By this I mean that PSTs are expected to know how to write objectives, design lesson plans, and write assessments. Even though they do not yet have methods knowledge, they still are expected to abstractly design lessons and create assessments that will measure student knowledge. As part of knowing structure, PSTs are expected to develop familiarity with technological teaching tools and understand how and when they can be used.

PSTs are also expected to be professional and exhibit teacher dispositions. They are expected to come on time to class, submit work by the deadlines, and collaborate with peers. They are expected to present their work using SAE, and to pay close attention to grammar and spelling. They need to demonstrate that they can reflect upon their experiences and describe a willingness to grow and change.

# Research question 2: How do teacher educators assess these middleprogram PSTs?

PSTs are assessed using a dual requirement at GU, the course grade and the core assignment. The core assignment is departmentally made and PSTs need to score a C or higher on it. This assignment, however, can be adapted by individual course instructors and is graded by these instructors. This means that the weighting and focus of the core assignment can change, even though the general structure remains constant. From my analysis, it appeared that PSTs were

assessed both on the content provided in the core assignment submission, as well as graded on how well they adhered to the guidelines. The rubrics included elements that focused on professionalism, rule following, and SAE grammar. Thus, for the core, PSTs needed to both answer all the questions and present their material in a clear and prescribed manner.

The second requirement, the course grade, was determined using a myriad of components that were at the discretion of the course instructor. In general, the components of the grade included written assignments and class behavior. For graded assignments, PSTs needed to demonstrate that they could exhibit the taught skills (e.g. design a Jigsaw lesson, create a table to analyze student test results, or build a WebQuest). These assignments varied in length and weight. PSTs also needed to attend class regularly and turn in assignments on time, as points were deducted for late submissions. Some of the courses also graded in-class participation, requiring PSTs to be engaged in class and to interact with peers. Additionally, while it was not universal, some instructors also administered exams based on the taught knowledge. For this requirement, PSTs were assessed on their demonstrated understanding of the course material, as well as on their teacher dispositions.

# Research question 3: What are the tensions involved in these expectations and assessment decisions?

As grading and assessment are controversial topics, several tensions arose as I came to understand how PSTs are assessed in the middle of the program. Much of these tensions surrounded issues of fairness, as grading necessarily discriminates (although ideally only using the mathematical definition). Dr. Polaris's claim of, "If it's important enough to teach, I would like to know if they learned it" was not always possible (transcript). Because of the nature of middle program courses, there were many proxies for assessing teacher knowledge. PSTs were often graded on their degree of professionalism and rule-following, as these behaviors were easier to see. Furthermore, as middle program courses focused more on structure than substance, there needed to be a degree of leniency in how content was assessed. The **PSTs** had not yet taken methods courses, and so they designed lesson plans, assessments, and technology activities without really knowing how to conceptualize knowledge for students. Thus, the grading of these aspects focused on what they did know so far, which was the set-up. The degree to which measuring structure had predictively validity for use in the future classroom, however, was not made explicit. Issues of fairness were also present when one considered the grading scheme and how it varied from course to course and section to section. The target and the focus moved, and consequently what mattered and was assessed changed. These tensions, described in more detail earlier in this chapter, all contributed to the tenuous relationship between teaching and grading. As I mentioned in the introduction, "the combination of enthusiastic support and strong disapproval [for assessment] has a long history" (Linn, p. 29).

## **Concluding Thoughts**

Through this dissertation, I was able to examine the ways that PSTs are assessed in the middle of their program. As my first foray into deeply qualitative research, I found the experience meaningful. I maintained my roots as a quantitative researcher as I organized my data in excel and ran some calculations. Together, these methods led me to develop myself as mixed-methods researcher and gave me a myriad of experiences that I will bring with me into my future.

As I looked closely at three cases within one university, I was able to consider the nuances and reasons for the assessment decisions I encountered. I uncovered what could be learned about a course when only certain materials were considered, and I was able to build matrices to look at courses more concretely. Thus, I learned about what PSTs at GU were expected to learn, know,

and do. I learned how PSTs were assessed. Finally, I looked at tensions in assessment, both as the explicitly showed themselves in the courses, as well as from a more philosophical standpoint. Assessment is a big category, and assessing fairly is important. Thus, through my dissertation I have contributed to the field of teacher preparation assessment. I am hopeful that what I wrote about here can be used to further the understanding of what happens in the middle of a teacher preparation program and why. APPENDICES

## Appendix A

Table A.1 Detailed timeline of testing from mid-1800s through 1990s				
Year	Fact	Source		
mid 1800s	Edouard Séguin used foam boards with cognitively impaired children	Boake (2002)		
1880s	Galton and Joseph Jacobs uses Digit Span Test	Boake (2002)		
1897-1927	Ralph Tyler says this period was full of many criticisms of testing	Haney (1981)		
1884	Galton's "anthropometric" measures administered at the International Health Exhibit in London	Boake (2002)		
1887	Jacobs details the origin of the test instructions for the Digit Span Test	Boake (2002)		
1887	Jacob's coins "prehension"	Boake (2002)		
1890s	Cattell adapts Galton's tests	Boake (2002)		
1890	James McKeen Cattell coined the term "mental test"	Boake (2002)		
1890	Joseph Rice administers spelling surveys, which is sometimes seen as the beginning of standardized testing in the United States	Haney (1981)		
1895	Binet and Victor Henri review "differential psychology" tests that later make it into the intelligence test battery	Boake (2002)		
1900	Creation of the Substitution Test	Boake (2002)		
1905	Binet-Simon intelligence scale published (and developed for Paris school children)	Boake (2002)		
1908	Simon and Binet update their test to have age levels	Boake (2002)		
1910s	Ellis Island psychologists perform intelligence testing on immigrants	Boake (2002)		
1910	Bosney develops arithmetic reasoning items	Boake (2002)		
1911	Healy and Fernald develop form board items	Boake (2002)		
1911	Using non-verbal tasks to measure intelligence is coined "performance" testing	Boake (2002)		
1916	Lewis Terman extends Binet-Simon into adulthood and turn "mental age" into "intelligence quotient"	Haney (1981)		
1917	Pintner-Paterson Performance Scale for assessing children with hearing impairments	Boake (2002)		
Just before WWI	Robert Yerkes and James Bridges change Binet-Simon scale from years to points.		Boake (2002)	
WWI	United States uses Examinations Alpha and Beta to measure intelligence of military recruits.	Boake (2002)		
WWI	Army Individual Examination developed for English-speaking militarecruits, but never administered because it failed the standardization phase. However, it was majorly influential for the Wechsler intelligence and memory scales.	ary 1	Boake (2002)	

Table A.1	(Continued)				
WWI	Army Performance Scale designed for recruits who did poorly on the Army group examinations or were not strong English speakers.	Boake (2002)			
1917-1919	Over 1.7 millions military recruits administered Alpha or Beta	Boake (2002)			
1917	Wechsler begins working to score Alpha examinations	Boake (2002)			
1918-1922	Wechsler studies with Charles Spearman, Karl Pearson, Henri Piéron and others overseas	Boake (2002)			
1919	News article suggests that the Army Trade Tests should be used in vocational schools	Haney (1981)			
1921	<i>Education Review</i> publishes an article on using intelligence tests for college admissions	Haney (1981)			
1921	School and Society publishes article on using army intelligence tests In high schools	Haney (1981)			
1922	Walter Lippmann and Terman begin long, published debate about testing and its worth	Haney (1981)			
1926	SAT is introduced as part of college admissions	Haney (1981)			
1927-1938	Ralph Tyler says this period responded to the criticism and extended what was tested	Haney (1981)			
WWII	Army still uses tests, but receives much less attention for it	Haney (1981)			
1930s	Oscar Burros begins career as bibliographer of testing, starting the Mental Measurement Yearbook	Haney (1981)			
1937	Monroe notes three trends in testing: emphasis on validity, focus on direct (instead of indirect) measurement, increased attention on essay tests	Haney (1981)			
1950s	Focus on tracking and selection	Linn (2000)			
1957	Sputnik is launched	Haney (1981)			
1958	National Defense Education Act (NDEA) provided financial assistance for schools to administer testing	Haney (1981)			
1960s	Tests used for program accountability	Linn (2000)			
Table A.1	Table A.1 (Continued)				
1960s	Public attention to personality testing	Haney (1981)			
1960s	More privacy for testers regarding their personality tests	Haney (1981)			
1960	New York Times posts article on "What the tests do not test"	Haney (1981)			

1069	Banosh Hoffmann publishes The Tyrappy of Testing	Hanov
1902	Danesh Honmann publishes The Tytanny of Testing	(1981)
1970s	Tests used for minimum competency testing	Linn (2000)
1970s	Controversy about IQ tests and many (racist and classist) arguments suggesting that IQ is genetic	Haney (1981)
1970s	Truth-in-testing legislation	Haney (1981)
1971	Richard Herrnstein argues that because of the heritability of IQ, the US will soon have a caste system.	Haney (1981)
Late 1970s	SAT scores drop and College Board and ETS study the problem, concluding that society and lack of motivation is the problem (ignoring that maybe the test is just no longer applicable in its current form)	Haney (1981)
1979	Debra P. v. Turlington questions the legitimacy of standardized literacy tests in Florida as requirement for high school diploma	Haney (1981)
1979	Barbara Lerner, director of National Academy of Sciences Committee on Ability Testing, declares that there is "war on testing." She claims that the National Education Association is fighting tests because it reveals what poor educating they have been doing	Haney (1981)
1976-1980	States begin adopting "minimum competency testing" for awarding high school diplomas	Haney (1981)
1980s	Tests used for district and school accountability	Linn (2000)
1980	Ralph Nader goes on the Johnny Carson show to condemn ETS	Haney (1981)
1980	New York's Truth-in-testing law required that test questions and answers be released within 30 days of score release	Haney (1981)
1980	MCAT gets legal exemption from New York's law	Haney (1981)
1980	George Hanford, President of College Board, declares there is a "war on testing"	Haney (1981)
1990s	Tests used for standards accountability	Linn (2000)

## Appendix B

## Interview Protocol

- 1. How would you describe the purpose of your course? What is included/excluded in the scope of this course?
- 2. What do you expect students to already know when they enter your course?
- 3. What do you expect your students to be able to do/know after they leave this course?
  - a. What can you claim about your students after they finish this course?
- 4. How do you design your curriculum and lesson plans for the semester?
  - a. What choices are yours to make and what comes from others?
  - b. How do you choose what to add or cut?
- 5. How do you grade students in this course?
  - a. What is graded and how do you weight? Can you show me?
  - b. Are there any aspects of the course that you do not assess? What and why?
- 6. How do you decide to grade in this way? What influences your grading decisions?
  - a. What influences your grading decisions? Do you have control over the assessments in your course?
  - b. How is your curriculum affected by your assessment choices and vice versa?
- 7. What does it mean to be an A student in this course? What must a student do or know to be an A student?
- 8. What would constitute a student failing your course? What is the minimum that they can do and still pass?
- 9. How do you feel that the grading in your course matches what other instructors do?
  - a. Do you think this difference/similarity is important?

- b. How do you decide how and when you will do to match/differ from other instructors?
- 10. Would an A in your section be the same as an A in another?
  - a. Does this matter?
    - i. To you?
    - ii. To your students?
    - iii. To what the course claims to teach?
- 11. If you teach this course again:
  - a. Will you want it to be more similar or different to the sections taught by your colleagues?
  - b. Would you change any assessments? Which and why?
- 12. (Added during first interview) What is your conception of fairness when it comes to assessment?

REFERENCES

## REFERENCES

- Al-Rawashdeh, A., Ivory, G., & Writer, J. H. (2017). Evaluating the Dispositions of Teacher Education Candidates: A Place for Self-Assessment. *Journal of Educational and Psychological Studies [JEPS]*, 11(4), 749-761.
- Ananiadou, K. and M. Claro (2009), "21st Century Skills and Competences for New Millennium Learners in OECD Countries", *OECD Education Working Papers*, No. 41, OECD Publishing. http://dx.doi.org/10.1787/218525261154. Retrieved from <u>http://repositorio.minedu.gob.pe/bitstream/handle/123456789/2529/21st%20Century%20S</u> <u>kills%20and%20Competences%20for%20New%20Millennium%20Learners%20in%20OE</u> <u>CD%20Countries.pdf?sequence=1&isAllowed=y</u>
- Anyon, J. (1980). Social class and the hidden curriculum of work. Journal of education, 67-92.
- Archbald, D. A., & Newmann, F. M. (1988). Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School. Reston, VA: National Association of Secondary School Principals.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational researcher*, *36*(5), 258-267.
- An, H., Shin, S., & Lim, K. (2009). The effects of different instructor facilitation approaches on students' interactions during asynchronous online discussions. *Computers & Education*, 53(2009), 749-760.
- Baker, Eva L., Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard.
  "Problems with the Use of Student Test Scores to Evaluate Teachers. EPI Briefing Paper# 278." *Economic Policy Institute* (2010).
- Ball, L. C., Cribbie, R. A., & Steele, J. R. (2013). Beyond Gender Differences: Using Tests of Equivalence to Evaluate Gender Similarities. *Psychology of Women Quarterly*, 37(2), 147-154. DOI: 10.1177/0361684313480483
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of teacher education, 59*(5), 389-407.
- Beidleman, D. C., & Cole, C. L. (1991). Scholastic aptitude test gender gap. American Secondary Education, 19(2), 2-5. Retrieved from https://search.proquest.com/docview/1297776663?accountid=12598
- Benbow, C.P., & Wolins, L. (1996). Utility of out-of-level testing for gifted 7th and 8th graders using SAT-M: An examination of item bias. In C.P. Benbow & D. Lubinski

(Eds.), *Intellectual talent: Psychometric and social issues* (pp. 333-346). Baltimore, MD: Johns Hopkins University Press.

Berkeley, G. (1972). The principles of human knowledge. Collins.

- Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41(3), 287-302.
- Borko, H., Liston, D., & Whitcomb, J. A. (2007). Apples and fishes: The debate over dispositions in teacher education. *Journal of Teacher Education*, *58*(5), 359-364. doi:10.1177/0022487107309977
- Bourdieu, P. (1973). Cultural reproduction and social reproduction. Retrieved from http://edu301s2011.files.wordpress.com/2011/02/cultural-reproduction-and-socialreproduction.pdf
- Bourdieu, P. (1971). Reproduction culturelle et reproduction sociale. Social science information/information sur les sciences sociales, 10(2), 45.
- Brabeck, M. M., Dwyer, C. A., Geisinger, K. F., Marx, R. W., Noell, G. H., Pianta, R. C., . . . Worrell, F. C. (2016). Assessing the assessments of teacher preparation. *Theory into Practice*, 55(2), 160-167. doi: 10.1080/00405841.2015.1036667
- Breed, F., & Breslich, E. (1922). Intelligence tests and the classification of pupils II. The School Review, 30(3), 210-226. Retrieved from http://www.jstor.org/stable/1078544
- Carrol, N. & Burke, M. (2010). Learning effectiveness using different teaching modalities. *American Journal of Business Education, 3*(12), 65-76.
- Carter, N. (2003). Convergence or Divergence: Alignment of Standards, Assessment, and Issues of Diversity. AACTE Publications, 1307 New York Avenue, NW, Suite 300, Washington, DC 20005-4701.
- Clarke, M. M., Madaus, G. F., Horn, C. L., & Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, 32(2), 159-181.
- Cochran-Smith, M., & Maria Villegas, A. (2015). Studying teacher preparation: The questions that drive research. *European Educational Research Journal*, 14(5), 379-394.
- Committee, O. T. S. O. T., & National, R. C. (2010). Preparing teachers: Building evidence for sound policy. Retrieved from https://ebookcentral-proquest-com-proxy1-cl-msuedu.proxy2.cl.msu.edu
- Council for the Accreditation of Educator Preparation. (2016). 2016 CAEP Standards for Advanced Programs. Retrieved from

http://www.caepnet.org/~/media/Files/caep/accreditation/caep-adv-program-standards-one-pager.pdfPla=en.

- Creswell, J.W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage, Thousand Oaks, CA.
- Croft, S. J., Roberts, M. A., & Stenhouse, V. L. (2015). The perfect storm of education reform: High-stakes testing and teacher evaluation. *Social Justice*, 70-92.
- Darling-Hammond, L. (2010). Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching. *Center for American Progress*.
- Darling-Hammond, L. (2010). Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching (pp. 1-27, Rep.). Washington, DC: Center for American Progress.
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, 53(4), 286-302.
- Descartes, R. (1984). Meditations IV. John Cottingham et al.
- Faber, R. (2008). Gender bias in the Trends in Mathematics and Science Study 2003 (TIMSS) for Canadian students. (MR46561 M.Ed.), Brock University (Canada), Ann Arbor. Retrieved from http://ezproxy.msu.edu/login?url=http://search.proquest.com/docview/304819495?accounti d=12598 ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Full Text database.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, 12(2), 219-245.
- Forzani, F. M. (2011). *The Work of Reform in Teacher Education* (Unpublished doctoral dissertation). University of Michigan, Ann Arbor, Michigan.
- Gallanger, A., Levin, J., & Cahalan, C. (2002). Cognitive patterns of gender differences on mathematics admissions tests. *ETS Report Series, 2002*(2). Princeton, NJ: Educational Testing Service.
- Ginsberg, R., & Kingston, N. (2014). Caught in a vise: The challenges facing teacher preparation in an era of accountability. *Teachers College Record*, *116*(1), n1.
- Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education*, *68*(4), 377-393.

- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. Economics of Education Review, 34, 29-44.
- Goldin, C., & Rouse, C. (1997). Orchestrating impartiality: The impact of "blind" auditions on female musicians (No. w5903). National bureau of economic research.
- Gorski, P. C. (2009). What we're teaching teachers: An analysis of multicultural teacher education coursework syllabi. *Teaching and Teacher Education*, *25*(2), 309-318.
- Greenberg, A. C. (2010) Fighting Bias with Statistics: Detecting Gender Differences in Responses to Items on a Preschool Science Assessment. (Doctor of Philosophy), University of Miami. Retrieved from http://ezproxy.msu.edu/login?url=http://search.proquest.com/docview/870283125?accounti d=12598 ERIC database.
- Haertel, E., & Calfee, R. (1983). School Achievement: Thinking about What to Test. Journal of Educational Measurement, 20(2), 119-132. Retrieved from http://www.jstor.org/stable/1434660
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*(1), 17-27.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice itemwriting guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333. doi:10.1207/S15324818AME1503\_5
- Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. American Psychologist, 52(10), 1091-1102. doi: http://dx.doi.org/10.1037/0003-066X.52.10.1091
- Haney, W. (1981). Validity, vaudeville, and values: A short history of social concerns over standardized testing. *American Psychologist, 36*(10), 1021-1034.
- Heafner, T., McIntyre, E., & Spooner, M. (2014). The CAEP standards and research on educator preparation programs: Linking clinical partnerships with program impact. *Peabody Journal* of Education, 89(4), 516-532.
- Huang, C. J., & Oga-baldwin, W. (2015). Assessing outcomes of teacher education: Quantitative case studies from individual Taiwanese and Japanese teacher training institutions. *The Asia Pacific Education Researcher, 24*(4), 579-589. doi: http://dx.doi.org.proxy1.cl.msu.edu/10.1007/s40299-014-0203-4
- Kincheloe, J. L. (2008). *Critical Pedagogy Primer* (2<sup>nd</sup> ed). New York, NY: Peter Lang.

- Keramidas, C. G. (2012). Are undergraduate students ready for online learning? A comparison of online and face-to-face sections of a course. *Rural Special Education Quarterly*, 31(4), 25-32.
- Kock, N., Verville, J., & Garza, V. (2007). Media naturalness and online learning: Findings supporting both the significant-and no-significant-difference perspectives. *Decision Sciences Journal of Innovative Education*, 5(2), 333-355.
- Kraft, N. P. (2001). Standards in Teacher Education: A Critical Analysis of NCATE, INTASC, and NBPTS (A Conceptual Paper/Review of the Research).
- Lancaster, J. W., Wong, A., & Roberts, S. J. (2012). 'Tech'versus 'Talk': A comparison study of two different lecture styles within a Master of Science nurse practitioner course. Nurse education today, 32(5), e14-e18.
- Lee, H. J. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching* and teacher education, 21(6), 699-715.
- Loewen, J. W., Rosser, P., Katzman, J., & Women's Educational Equity Act, P. (1988). Gender bias in SAT items. (Report No. TM 011 653). New Orleans, LA. Government Document: American Educational Research Association.
- Luo, Y., Pan, R., Choi, J. H., Mellish, L., & Strobel, J. (2011). Why choose online learning: Relationship of existing factors and chronobiology. *Journal of educational computing research*, 45(4), 379-397.
- Mursell, J. L. (1939, Dec 01). Mental Testing: A protest. *Harper's Magazine*, 180, 526-534. Retrieved from http://ezproxy.msu.edu.proxy1.cl.msu.edu/login?url=https://searchproquest-com.proxy1.cl.msu.edu/docview/1301529069?accountid=12598
- National Research Council. (2001). Testing Teacher Candidates: The Role of Licensure Tests in Improving Teacher Quality. Washington, DC: The National Academies Press. doi: 10.17226/10090.
- Navarro, C. (1989). Why Do Women Have Lower Average SAT-Math Scores than Men? (Report No. TM 014 018). San Francisco, CA: American Education Research Association.
- Norris, J. M. (2013). Some challenges in assessment for teacher licensure, program accreditation, and educational reform. *The Modern Language Journal*, *97*(2), 554-560.
- O'Connor, M. C. (1992). "Rethinking aptitude, achievement, and instruction: Cognitive science research and the framing of assessment." In B.R. Gifford & M. C. O'Connor (Eds.), *Changing Assessments* (pp. 9-35). New York, NY: Springer Science+Business Media.
- Patton, Michael Quinn (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, CA: Sage.

- Pecheone, R. L., Pigg, M. J., Chung, R. R., & Souviney, R. J. (2005). Performance assessment and electronic portfolios: Their effect on teacher learning and education. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 78*(4), 164-176.
- Perie, M., Marion, S., & Gong, B. (2009). "Moving toward a comprehensive assessment system: A framework for considering interim assessments." *Educational Measurement: Issues and Practice, 28*(3), 5-13.
- Pinter, R. (1923). Intelligence Testing. New York, NY: Henry Holt and Company.
- Raykov, T., & Marcoulides, G. A. (2011). Introduction to psychometric theory. Routledge.
- Ryan, B., Scapens, R. W., & Theobald, M. (2002). *Research method and methodology in finance and accounting.* Padstow, United Kingdom: TJ Digital.
- Rosser, P. (1989). *The SAT gender gap: Identifying the causes.* Washington, DC: Center for Women Policy Studies.
- Scates, D. E. (1950). The good teacher: Establishing criteria for identification. *Journal of Teacher Education*, 1(2), 137-141.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.
- St. George, D. (2014, June 28). Steep failure rate on Algebra I exams in Montgomery leads to mass recalculation. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/steep-rates-of-failure-on-algebra-exams-inmontgomery-lead-to-mass-recalculation/2014/06/28/f75eeaf2-fe4c-11e3-932c-0a55b81f48ce\_story.html?utm\_term=.270aa634acba on 7 October 2017.
- Stake, R. E. (1995). The art of case study research. Sage.
- Suzuka, K., Sleep, L., Ball, D. L., Bass, H., Lewis, J., & Thames, M. (2009). Designing and using tasks to teach mathematical knowledge for teaching. Scholarly practices and inquiry in the preparation of mathematics teachers, 7-24.
- Taylor, R. L., & Wasicsko, M. M. (2000, November). The dispositions to teach. In annual meeting of the Southern Region Association of Teacher Educators (SRATE) Conference, Lexington, KY. Retrieved May (Vol. 21, p. 2010).
- "The Word Neutral Man's Burden The Colbert Report." *The Colbert Report*, episode 586, Comedy Central, 16 July 2009, www.cc.com/video-clips/tt0y6c/the-colbert-report-the-word---neutral-man-s-burden.

- Thompson, J. R., Klass, P. H., & Fulk, B. M. (2012). Comparing online and face-to-face presentation of course content in an introductory special education course. *Teacher Education and Special Education*, 35(3), 228-242.
- Tulsky, D. S., Saklofske, D. H., & Ricker, J. (2003). Historical Overview of Intelligence and Memory: Factors Influencing the Wechsler Scales. *Clinical interpretation of the WAIS-III* and WMS-III, 7.
- Villegas, A. M. (2007). Dispositions in teacher education: A look at social justice. *Journal of Teacher Education, 58*(5), 370-380.
- Walker, H. (1925). What the tests do not test. *The Mathematics Teacher, 18*(1), 46-53. Retrieved from http://www.jstor.org/stable/27950686
- Walsh, K., Glaser, D., & Wilcox, D. D. (2006). What education schools aren't teaching about reading and what elementary teachers aren't learning. *National Council on Teacher Quality.*
- Wiest, L. R. (2008). Problem-solving support for English language learners. *Teaching Children Mathematics*, 14(8), 479.
- Wiggins, G., & McTighe, J. (2001). What is backward design?. Understanding by design, 7-19.
- Wineburg, M. S. (2006). Evidence in teacher preparation: Establishing a framework for accountability. *Journal of Teacher Education*, 57(1), 51-64. doi: 10.1177/0022487105284475
- Worcester, D. A., & Corey, S. M. (1936). In criticism of the Detroit tests of learning aptitude. *Journal of Educational Psychology*, 27(4), 258-262. doi:http://dx.doi.org.proxy2.cl.msu.edu/10.1037/h0060644