

Decisions and Tensions: Summative Assessments in PBL Advanced Placement Classes

Susan Elizabeth Cooper

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Dr. Susan B. Nolen, Chair

Dr. Kenneth M. Zeichner

Dr. Catherine S. Taylor

Program Authorized to Offer Degree:

College of Education

©Copyright 2017
Susan Elizabeth Cooper

University of Washington

Abstract

Decisions and Tensions: Summative Assessments in PBL Advanced Placement Classes

Susan Elizabeth Cooper

Chair of the Supervisory Committee:
Professor Susan B. Nolen
College of Education

This study examines how teachers navigate tensions between communicating expectations for college level work and student motivation as they determined summative grades in an AP Environmental Science course. Semi-structured interviews and a think-aloud protocol were conducted while teachers in poverty-impacted urban high schools determined final semester grades. Analysis centered on the ways these teachers negotiate the tensions between college preparation (rigor) and keeping students engaged (rather than discouraged) as they determine report-card grades. Teachers in the study showed differences in how they negotiated these decisions and tensions as they grappled with delivering high level content to students choosing to take on the challenge of an AP Science course. This study also revealed the tensions and negotiations involved in the strategic decisions teachers must make while balancing the rigor of a college curriculum with providing equity of access to future learning opportunities in higher education for poverty impacted students while taking on district changes in policy and new mandates regarding assessment. Findings show that given multiple competing inputs from

various communities of practice, teachers structure their decision making in what they believe is in the best interest of their students so that they can go on to higher education opportunities. Looking at the multiple inputs from others also suggests the onus is on the teacher to broker what the grades they give mean to multiple groups invested in students' grades for a variety of reasons. Implications for how to support teachers' summative assessment practices for those who have open access to their AP courses are discussed.

Table of Contents

| | |
|---|------|
| List of Figures..... | vi |
| List of Tables..... | vii |
| Acknowledgements..... | viii |
| Dedication..... | xi |
| CHAPTER 1: Introduction..... | 1 |
| CHAPTER 2: Review of the Literature, Theoretical Frameworks, and Context of the Study..... | 4 |
| 2.1 Review of the Literature on Grading Practices..... | 4 |
| 2.2 Theoretical and Frameworks..... | 9 |
| 2.2.1 Grades as Boundary Objects..... | 9 |
| 2.2.2 Teacher Decision Making..... | 13 |
| 2.2.3 Research Questions..... | 16 |
| 2.2.4 Context of the Study..... | 16 |
| 2.2.4.1 Assessment in the Project-Based Learning Classroom..... | 18 |
| CHAPTER 3: Methodology..... | 20 |
| 3.1 Participants and Setting..... | 20 |
| 3.2 Data Collection..... | 22 |
| 3.3 Researcher Positionality..... | 23 |
| 3.4 Data Analysis..... | 24 |
| CHAPTER 4: Findings..... | 26 |
| 4.1 Summary of Findings..... | 26 |

| | |
|--|----|
| 4.2 Claim 1: AP teachers acted as brokers to help move their students from the world of high school to college so that their students were prepared for college..... | 26 |
| 4.3 Claim 2: AP teachers experienced tension as the acted as brokers between the worlds of high school and college while supporting student motivation and persistence in opportunities to learn..... | 28 |
| 4.4 Claim 3: AP teachers felt tensions between their roles as brokers between high school and college and the grading practices mandated in the district’s new policy of Standards-Referenced Grading (SRG)..... | 36 |
| 4.5 Findings Conclusion..... | 38 |
| CHAPTER 5: Discussion and Conclusion..... | 42 |
| 5.1 Discussion..... | 42 |
| 5.2 Conclusion..... | 46 |
| 5.2.1 Limitations..... | 47 |
| 5.2.2 Implications..... | 47 |
| References..... | 50 |
| Appendix 1..... | 56 |
| Appendix 2..... | 59 |

List of Figures

| | |
|---|----|
| Figure 1. Grades as boundary objects..... | 13 |
| Figure 2. Teachers as Reflective Decision Makers..... | 15 |
| Figure 3. Summative grading as a “balancing act”..... | 30 |

List of Tables

| | |
|---|----|
| Table 1. Demographic Characteristics of the Schools (%) – Part 1..... | 21 |
| Table 2. Demographic Characteristics of the Schools (%) – Part 2..... | 21 |

Acknowledgements

I wish to thank my supervisory committee for their patience, kindness, wisdom, and generosity. My advisor and mentor, Dr. Susan Nolen, has picked me up and dusted me off almost as many times as my parents did while I was growing up. She is one of the most generous people I know. She has always respected my unique funds of knowledge and has encouraged me every step of the way. Dr. Kenneth Zeichner inspired me to be proud of the public school teacher and parent that I am. His classes on teacher education challenged me to critically look at teacher education around the world and were instrumental in solidifying the notion that I will always carry with me – that being a teacher is the noblest and most humbling of professions. Dr. Catherine Taylor took a chance on me a few years back and handed me her Assessment for Teacher Candidates course when she went on to other endeavors. Warm and kind, she shared her knowledge and all her resources with me. Thank you for your mentorship. Dr. Stewart Tolnay welcomed me into his sociology class where I read the best books and engaged with and listened to some of the wisest young people engaged in social justice reform. Thank you for taking an interest in my work and for being a wonderful listener.

Thank you to the National Science Foundation. Work on which this dissertation is based was supported, in part, by a grant from the National Science Foundation, DRL 1251595, to Susan B. Nolen. All statements herein are the author's and do not represent the views of the NSF.

To my learning sciences cohort, Dr. Kendall Becherer, Dr. Sarah Evans, Dr. Theresa Horstman, Dr. Déana Scipio, Dr. Gavin Tierney, Dr. Ellie Walsh, and Dr. Sarah Ward: You have been the best part of these eight years. Despite how different we all are, never did our loyalty to one another falter. Thank you for your unconditional love and support. I will always treasure the memories made with all of you.

When I started this journey, I was still a full-time classroom teacher with four teenagers at home. My husband, David, did the carpooling, cooking, and cleaning so I could study. David, I don't know how you do it, but you always play me the right song just when I need it most. And to our now five children, McCoy, Zelda and Hussein, Emily and Katie: I am proud of the young adults you have become. You are beautiful, wise, and good people. I am so proud of my quirky, blended, multicultural family.

To all my students over the past 25 years: Thank you for giving me the honor of being a small part of your lives. I'm still always amazed to receive a letter from a former student telling me the impact I had on his or her life. Little do you all know how much you taught me. It was because of all of you that I was a successful teacher. You challenged me, you were funny, and you taught me the importance of a safe, family-style classroom community. I am continuously humbled when I see the challenging work students do in public schools every day across the country.

And to my father, mother, sister, and brother: Your love and belief in me has always been unwavering. Thank you for letting me live an unconventional life – I know it wasn't always easy for you, but you never stopped loving me as I pursued authenticity in my life. Thank you for giving my children the best grandparents, aunt, and uncle a kid could ever hope for. They really lucked out having you to love and support them.

I am proud to say I am the oldest child of Kent and Elizabeth Boswell. From my father I inherited his humor, his crankiness, and his joy of working with young people. I have also learned the true meaning of love as I watch him care for my mother - always making sure she is treated with the utmost dignity. From my mother I received my love of reading and learning. She was also a ferocious advocate with total faith in me. These two Iowans married young, and went

on to raise three healthy, happy children in 1970's Los Angeles. Thank you for the wonderful childhood. Mom, you are the wisest woman I know. As long as you keep tucking my messy hair behind my ears, I am certain that you know who I am and what I have accomplished.

Dedication

For my mother, Elizabeth Boswell

CHAPTER 1: Introduction

In American education, grades are considered a time-honored and necessary tool for supporting the goals of learning and preparing students for life in the “real world” after formal schooling ends. Grades serve as communication to families and other community members (i.e., school and district personnel, policymakers, etc.), about the quality of education their children are receiving. Grades are often gate-keepers - determining courses of study, scholarships, and opportunities for higher education. With very little explanation, a grade of an A, a C, or an F is assumed to represent a student’s level of achievement (excellent, good, fair, poor, failure) in a given course of study. Even as student demographics in American schools change along with pedagogical practices, grades’ meanings, however, have persisted through the years. Grades across contexts imply similar and additional meanings (i.e., “The Easy A” or “Grading on the Curve”) that are shared. However, in Advanced Placement (AP) courses, grades are presumed to represent an assessment of students’ achievement at the college level.

With rigorous coursework and alignment to undergraduate college classes, these AP courses in high schools have historically been competitive and selective and often considered the “gold standard” of American education and college preparedness (Boss et al., 2011). However, issues of equity and access have led many school districts to remove entrance requirements to encourage more students to enroll in advanced placement classes (Riley, 2005; Schneider, 2009). At the same time, however, the number of students who fail AP exams is growing (Lewin, 2010). Given the diversity of the students in these newer, inclusive AP classes, how can the learning needs of our current AP students be accommodated? In particular, how do students in poverty-impacted schools gain deep conceptual learning in an advanced placement class while negotiating the goal of the College Board to prepare students for the challenges of college work?

Factors such as increase in work load and course readings as they prepare for a high stakes test with the tangible reward of college credit all have to be negotiated in a new context where students do not necessarily have the time and means to prepare in ways that students in more well-resourced schools do. Although expanding access to AP classes allows for essential moves towards diversity and equity, negotiating the different tensions related to the quality of learning in AP coursework is also important.

For many AP students, grades are very important. Often college-bound, students rely on grades and work hard to achieve grades to gain access to universities. Besides receiving college credit, students expect AP courses to be as rigorous and as similar in structure to college classes as possible. Grades in their AP courses may communicate to students their likelihood of success on both the AP exam and in college coursework. However, the National Research Council (2002) notes that opening traditional AP courses to a diversity of students may not lead to their successful performance on AP exams. Teachers in open access classes¹ have to negotiate many different tensions as they grade their students. The grades given in an AP class serve multiple roles and communicate a student's progress to the student, parents, administrators, and the public. Have their students learned the content? Will they be ready for the rigors of college? Will their students be successful enough in the AP class so that they persist and stay motivated to continue as learner? In addition, there is the added pressure on teachers to prepare students for an AP exam that gives additional information about the student's readiness for college and also provides a tangible reward for passing: college credit.

The aim of this paper is to look at the grading practices of five AP Environmental

¹ "Open Access AP Classes" in this context refers to no testing requirement for placement into the AP class. The only prerequisite necessary to take AP Environmental Science in this study is the completion of a biology course.

Science teachers who teach in poverty-impacted schools using Project-Based Learning, an ambitious form of pedagogy. Given that they are not using the traditional AP curriculum (lectures and textbook readings and assignments) and that they are held accountable to many different communities (i.e., colleges, the College Board, students, families, and the school communities) interested in the grades of AP students, how do teachers go about the everyday task of assigning grades with so many stakeholders' competing uses of grades?

CHAPTER 2: Review of the Literature, Theoretical Frameworks, and Context of the Study

Review of the Literature

In general, there is a paucity of literature regarding teacher grading policies and summative assessments and even less literature surrounding the grading practices of AP teachers. What does exist in the literature, beginning in the late 1980's up to now, is a growing mountain of literature regarding the effectiveness of formative and authentic assessments. As high-stakes testing accountability in the era of No Child Left Behind affected many districts, schools, teachers, and families, researchers sought to show the importance of classroom-based assessment and its role in scaffolding deep conceptual learning for students, informing better teaching practices for educators, and increasing engagement in learning (Allen, Ort, & Schmidt, 2009; Bass & Glaser, 2004; Black & Wiliam, 1998, 2009; Brookhart, 1997; Harlen & James, 2008; Hill & McNamara, 2011; Kirton, Hallam, Peffers, Robertson, & Stobart, 2005; Nicol & MacFarlane-Dick, 2006; Shepard, 2000; Stiggins, 2004; Wiliam, 2006). Teacher decision-making while grading and the results of those grading practices on students and other invested members of the community, such as school and district members, as well as local and more distant policymakers, have not been explored very much. This study specifically looks at how teachers come to make decisions about their students' grades.

Survey and descriptive studies of the grading practices of teachers report that despite district mandates and recommendations in the literature to give grades based on achievement only (Campbell, 2012; Jongsma, 1991; Marzano, 1996; O'Connor, 2009; Randall & Engelhard, 2010), many teachers use other factors such as effort, attendance, behavior, and attitude when

determining grades (Brookhart, 2011; Jongsma, 1991; Pope, Green, Johnson, & Mitchell, 2009; Randall & Engelhard, 2010). Absent from this literature is any insight on why teachers incorporate factors other than achievement. Brookhart (1991) found that even those teachers with assessment training and who have knowledge of measurement theory and principles of grading still struggle with the issue of grading on achievement alone. Unlike previous studies that have relied on teachers' self-report surveys of their practices, this study uses interview and think-aloud data as teachers grade, in the moment, to shed light on the considerations teachers may include in their decision-making when grading their students and communicating progress for end of semester grades.

The mismatch between grading policies and what teachers value as learning is the mysterious "black box" that Black & Wiliam (1998) refer to when they question what practices teachers engage in when assigning meaning and value to the work and learning their students do. Black and Wiliam define the "black box" as such:

...present policies in the U.S. and in many other countries seem to treat the classroom as a black box. Certain inputs from the outside - - pupils, teachers, other resources, management rules and requirements, parental anxieties, standards, tests with high stakes, and so on -- are fed into the box. Some outputs are supposed to follow: pupils who are more knowledgeable and competent, better test results, teachers who are reasonably satisfied, and so on. But what is happening inside the box? How can anyone be sure that a particular set of new inputs will produce better outputs if we don't at least study what happens inside? (p. 1)

Black and Wiliam go on to argue that the onus is often on the classroom teacher to make decisions despite a variety of inputs from outside the classroom with very little support from those providing the outside input. Until we study what goes on “inside the black box” we cannot determine if the inputs from the outside are helping us achieve the desirable outputs (such as mastery of content, increased critical thinking and motivation to learn, high rates of college-bound students, and low dropout rates) we want for our students (Black & Wiliam, 1998, p. 2). This study is structured to intentionally ask teachers what they are thinking about as they grade and what are they pulling from their “black box” - which is full of input from other groups they interact with – to make summative grading decisions.

Looking at the effect of teacher knowledge when given instruction in assessment and measurement, Brookhart’s 1993 study revealed that teachers still did not grade for achievement only. Her interpretation of why this may be:

When considering what consequences a grade will have for a student, the teacher functions as an advocate for the student. Concern about this function does not differ for those who have measurement instruction. This is consistent with the study’s hypothesis. Measurement instruction can be expected to clarify teachers’ concepts of the meaning of grades, but there is no reason to expect that measurement instruction will change thinking about values and social consequences. These thoughts are more closely related to the altruism that motivates people to enter teaching in the first place (Brookhart & Freeman, 1992). Child-centered orientations are powerful and pervasive among teachers. They may

help explain teachers' discomfort with the grading process (Barnes, 1985). If a teacher's first priority is to be an advocate for the student, concern about consequences to students may be expected to have more influence on grading practices than concern about interpretability. This priority may help explain why, in this study and others (Friedman & Manley, 1991; Stiggins & Conklin, 1992; Stiggins et al., 1989), teachers report considering effort as well as achievement when they assign grades (p. 140).

In the same article, Brookhart explores Bishop's (1992) notion that teachers need to see themselves more as advocates or "coaches" to their students as opposed to a "judge" who can determine if the student has adequately mastered content. That job of "judge", Bishop states in his article, cannot exist alongside the role of "coach" or advocate or mentor, and is more appropriate to give to external assessors. But this does not negate the need for the classroom teacher to be knowledgeable in assessment and measurement theories for "teachers would still need assessment information to identify levels of success and areas for improvement, in order to function as coaches" (p. 140).

With so much competing input entering the teachers' "black box", how do teachers manage it all when it comes time to summatively grade their students? Competing tensions and pressures abound as teachers somehow have to structure the information they are given, negotiate the needs of different communities of practice, and then make a decision that for many have important consequences. In particular, in an open access AP classroom, how do teachers structure their decision-making when grading to best help their students prepare for college *and* help their students see themselves as learners capable of moving on to higher education

possibilities? This study attempts to provide a structure of the teacher decision making process – in the moment - as the teachers completed their semester grades.

Along with competing input from other communities of practice, district and schools are dynamic places where curriculum and expectations are constantly changing to meet student or stakeholders' various needs. Like we find in this study, teachers are often having to negotiate some new policy or expectation (in this study's case it was a district move to a different type of grading system) that again has them figuring out again how to make sense of all of this input into the "black box" while at the same time grading their students for information about their learning and to motivate them to achieve.

Brookhart, Guskey, Bowers, McMillan, Smith, and Smith et al., in their 2016 study looking at 100 years of grading research (which included quantitative and qualitative studies alike, including survey and descriptive data) found that "compared to the number of studies about teachers' grading practices, relatively few studies focus directly on perceptual constructs as importance, meaning, value, attitudes, and beliefs" (p. 827). They found that looking at teacher grading perceptions in a multitude of various types of studies suggest four "clear and enduring" findings:

First, teachers idiosyncratically use a multitude of achievement and nonachievement factors in their grading practices to improve learning and motivation as well as document academic performance. Second, student effort is a key element in grading. Third, teachers advocate for students by helping them achieve high grades. Finally, teacher judgment is an essential part of fair and accurate grading (p. 828).

However, absent from these findings gleaned from over 100 years of grading studies is a clear concept of the “multidimensional construct of ‘success in school’” (Brookhart et al., 2016, p. 836) and how teachers’ instructional skills “merge” with their assessment skills to provide a picture of the quality of criteria teachers use to determine grades (p. 836). In the next section, I will look at the constructs of grades in various communities of practice hoping to understand the diverse meaning and values of grades along with the attitudes and beliefs many have regarding grades.

Theoretical and Conceptual Frameworks

Grades as Boundary Objects

Reporting grades is part of the important work teachers do. Besides reporting achievement, summative grades are used in many ways and can function as gate-keepers. Decisions such as “entry into advanced courses, promotion to the next grade, qualification for particular programs, scholarships, graduation, and college entry” (Taylor & Nolen, 2008, p. 369) are often made on the basis of summative grades. Grades do not just have local meanings – their meanings are negotiated with other worlds where grades can determine opportunities, accountability, and resources. As such, grades can take on other meanings and are used for purposes other than reporting the achievement of learning. In this way, grades serve as boundary objects (Star & Griesemer, 1989) among the different communities of practice who are invested in educational decision-making. Boundary objects are those objects that:

both inhabit several intersecting worlds and satisfy the informational requirements of each of them...[They are] both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a

common identity across sites. They are weakly structured in common use, and become strongly structured in individual site use. These objects may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds (Star & Greisemer, 1989, p. 393).

Grades, as assessment artifacts, are boundary objects in that they “have local meanings, within schools and classrooms, but at the same time, produce numbers and judgments that are highly consequential for other stakeholders in the educational system...[and] assessment practices...[are] particularly subject to multiple fronts of negotiation” (Nolen, Horn, Ward, & Childers, 2011, p. 89). Because grades function at the boundaries of different educational communities, Nolen et al. (2011) see assessment practices and tools as “potential bridges to communication but also as tools for learning and development” (p. 89).

In this study, the values, goals, and meanings of grades may not be the same for the different communities of practice in which they are used. As some of the literature points out (Briscoe, 1994; Brookhart, 1994; Campbell, 2012; Harrison et al., 2014; Thomas & Oldfather, 1997), classroom teachers often consider other factors when grading students because they may be concerned with the more interpersonal and immediate effects of grades as opposed to some of the gate-keeping aspects used by others. In short, different communities of practice may have competing ideas for the use of grades. The differences in use at the competing communities’ boundaries are often where negotiation of meaning and values can take place – a place where the

push and pull of each community's ideals can result in new understandings and learning. Wenger (1998) argued that "learning at the boundaries is necessary if communities of practice do not want to lose their dynamism and become stale" (as cited in Akkerman & Bakker, 2011, p. 136). Teachers work at the boundaries of different communities of practice daily. In this way, Akkerman and & Bakker see teachers as "brokers" or "boundary crossers." Citing the work of Fischer and Atkinson-Grosjean (2002), Akkerman & Bakker describe the work of "brokers":

On one hand, they have a task of "building bridges" between both worlds (Fischer & Atkinson-Grosjean, p. 463), being the means for connecting both sides. At the same time, however, these persons are held accountable in each world and must endure criticism; [for example] "by academics for being too aligned with industry, and by industry for being too academic" (p. 453)... On one hand they have a very rich and valuable position since they are the ones who can introduce elements of one practice into the other (cf. Wenger, 1998). On the other hand they face a difficult position because they are easily seen as being at the periphery, with the risk of never fully belonging to or being acknowledged as a participant in any one practice (p. 142).

Engeström, Engeström, and Kärkkäinen (1995) describe the work of brokers as those that "face the challenge of negotiating and combining ingredients from different contexts to achieve hybrid situations (p. 319). So although there is flexibility in each stakeholder's understanding of grades, it is on teachers to communicate, on the periphery of each community of practice they work with, to adjust to constraints (other worlds' values, meanings, and uses of grades) to

communicate effectively so that different communities of practice's needs are met. This may mean compromising teachers' own values, meanings, and uses of grades as they work to cross boundaries of other stakeholders' worlds in their work as brokers of the meanings of their students' grades.

For example, teachers must present grades to students and their families to communicate a student's level of achievement. They must also present grades to families and their school community to account for the learning successes (and failures) of their students. Grades are also used to communicate to students and their families anticipated success on the AP exam and their college career (especially in an AP classroom where the curriculum is designed to be "college level") and to help make choices about scholarships and other opportunities for study. Grades, communicating student achievement in a given community, are used by schools, districts, states, institutions of higher education, and policymakers to determine allocation of valuable and needed resources.

As an impetus for teacher learning, Ward, Nolen, & Horn (2011) term this inconsistency at the boundaries "productive friction" – the dissonance that results from the conflicting of two or more communities and which can "initiate positive changes in [the] use of high-leverage practices to improve student learning and understanding" (p. 15). Adopting the notion of productive friction may be helpful in understanding how different communities of practice can work together. This way grades as boundary objects can serve each groups' needs but also help mediate an AP classroom teachers' grading practices so that students understand how well they are doing in a rigorous course of instruction and, yet, stay motivated to continue high level content.

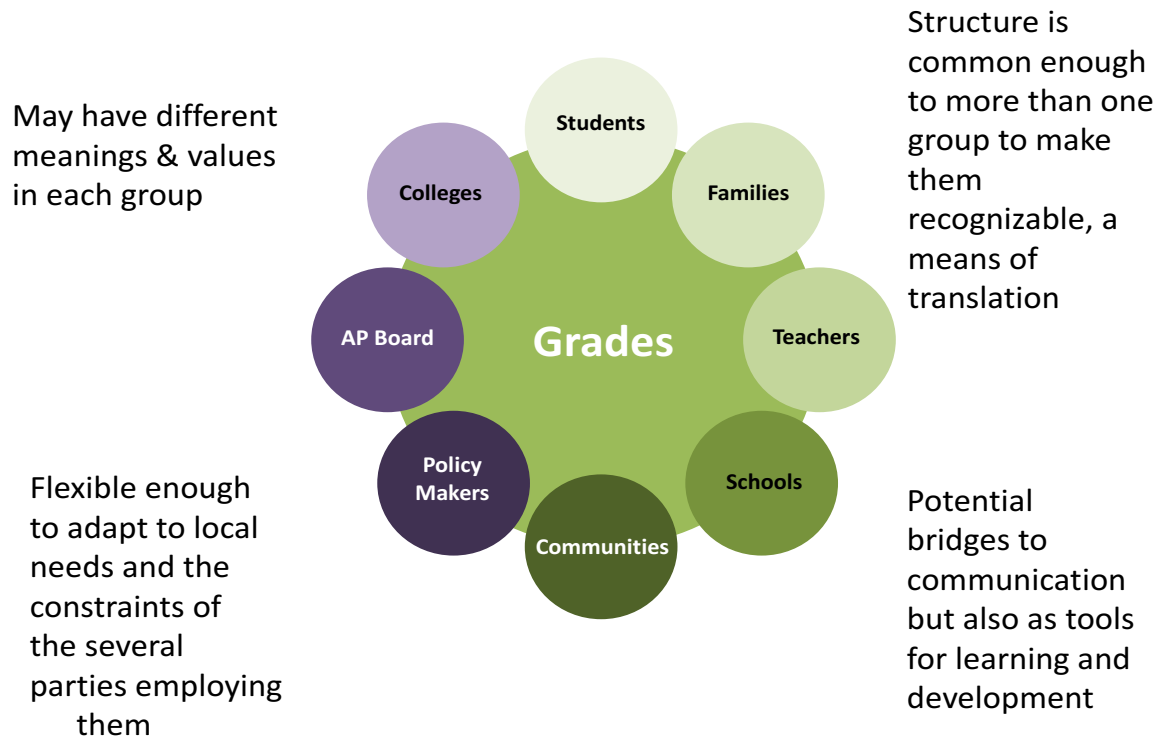


Figure 1. Grades as boundary objects

Teacher Decision Making

Brookhart (1991) noted that even those teachers with assessment training and who have knowledge of measurement theory and principles of grading still struggle with the issue of grading on achievement alone. Bishop (1992) suggest that this puts teacher in a difficult spot: they can be judges or they can be coaches for their students. Brookhart (1991, 1993), in her studies on validity showed that “in grading, teachers’ concerns over the many uses and the consequences of grade use sometimes outweigh their consideration of grade interpretation” (as cited in Cheng and Sun, 2015). Cheng and Sun continue:

As judges, they should base students' grades exclusively on achievement, giving priority to fairness, justice, and objectiveness of their judgment decision making. As coaches, they appeal to what is beneficial for students' development, building in considerations of many non-achievement factors, such as effort, encouragement, and improvement through grading (p. 215).

This duality and dichotomy of teachers' roles in grading adds to the tensions already experienced by AP teachers as they juggle getting students ready for the world of college and how they can grade their students to persist in further opportunities to learn. Bishop's 1992 article argued that "teachers cannot act as judges and coaches at the same time and suggested that teachers should give up the judging role to external assessment and focus on developing mentoring relationships with their students to fully function as coaches, mentors, or advocates for students in their own classroom assessment" (as cited in Cheng and Sun, 2015, p. 215). This is easier said than done as teachers are often brokering and compromising the meanings and values of grades as they work with many different communities of practice who see grades as used for other purposes. Looking back at Brookhart's 1993 study, we are reminded that teachers deal with many factors other than achievement even when they are skilled and educated in assessment and measurement. Coupling this with the dynamic nature of schools, in general, where new policies and changes in curricula and instructional practices are regular occurrences, teachers have many things to consider and negotiate when grading their students.

The following figure, taken from the St. John's University site on "Teachers as Reflective Decision Makers" illustrates how teachers may approach the grading process. Focusing on the "evaluating" part of grading I could see that these teachers were very much in the space of the

orange box: “Sift Options Through Criteria: Values, Professional Consideration, Constraints, and Personal Philosophy.”

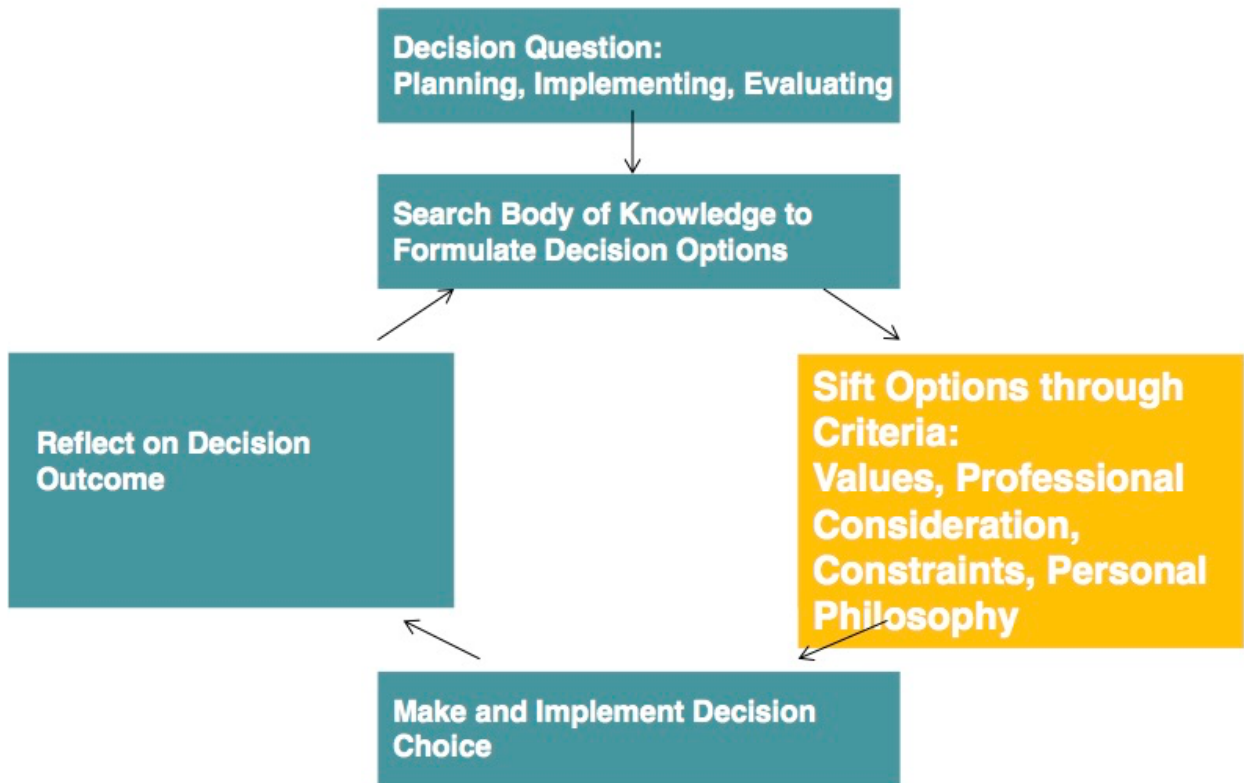


Figure 2. Teachers as Reflective Decision Makers (St. John’s University, 2017)

By asking teachers to think aloud while grading, I could join them in that orange box space as they were weighing the values, professional considerations, and constraints along with other inputs from different communities of practice they had to negotiate as they considered first semester grades for their AP Environmental Students. All five of these participating teachers,

although from different schools each with its own distinct contextual differences and values and meanings of grades, used the same APES Project-Based Learning curriculum and attended the same professional development sessions so that fidelity to the curriculum was maintained.

This study seeks to answer the following research questions:

1. What are the decision-making processes that teachers use to determine a summative grade for their students?
2. How do teachers negotiate the district-mandated grading policies, AP and College Board expectations, and other tensions from various groups in the context of open access AP courses when assigning summative grades?

Context of the Study

AP classroom teachers are charged with having to deliver a certain amount of course information in a certain amount of time. In order for their students to be prepared for a comprehensive exam, with tangible rewards (college credit) for passing, teachers are responsible for teaching a broad scope of information before the end of the school year. Students are expected to take in large amounts of information which contributes to the perception that AP courses are rigorous (Mellon, 2010). However, a report by the National Research Council in 2002 states, “the inclusion of too much accelerated content can prevent students from achieving the primary goal of advanced study: deep conceptual understanding of the content and unifying concepts of a discipline” (p. 1). Instead, the report recommends that AP programs “help students develop skills of inquiry, analysis, and problem solving so that they become superior learners” (p. 12).

The AP teachers in this grading study were recruited from a larger, parent study in which

AP teachers from different districts used project-based learning instructional techniques and tools to teach their AP courses. For this study, five AP Environmental Science (APES) teachers from the parent project participated. The teachers were from the same school district but each was from a high school. They agreed to be interviewed and videotaped as they produced end of semester grades to provide some insight into their thoughts during the grading process. (More about the participants and setting in the Methods section.)

In recent years, the district had greatly expanded AP offerings in all five schools and modified entrance requirements to create a more equitable AP opportunity for students. AP Environmental Science was one of the courses opened to all students who had completed the biology prerequisite in their neighborhood school. Open Access to AP courses comes with a set of challenges. As Boss et al. note:

Increasingly, school districts are lowering or removing AP entrance requirements and encouraging all students to tackle these “rigorous” courses. The motivation appears to be a combination of the new emphasis on “college readiness for all” plus increased attention to underserved urban schools with their greater number of students in or near poverty. The new thrust is that all students, in the name of equity, should have access to the gold standard... Although the number of exam takers is growing, the number of students who fail AP is growing, too. Some believe this is inevitable and worth the cost... (Boss et al., 2011, pp. 13-14).

Researchers and teachers in the parent project confronted this tension of opening AP courses to increasing numbers of diverse students by changing the curriculum and pedagogical approach

used in AP courses.

The approach used in the parent study was Project-Based Learning (PBL). In their book, *Powerful Learning: What We Know About Teaching for Understanding*, Darling-Hammond et al. (2008) argue that to prepare students for the skills they will need in the future, teachers need to engage their students in meaningful activities that allow students to apply knowledge and skills to real world problems. Instead of looking for right answers, students display powerful learning when they are involved in the process of inquiry, collaboration, and problem solving.

Assessment in the Project-based Learning AP Classroom

This pedagogical paradigm shift that is made in a PBL AP classroom exists in a context where success and learning in the course are still determined by a high-stakes test (the AP exam) before the end of the school year. Often teachers have to straddle the two worlds of PBL and AP – each with a different set of goals and expectations - and somehow communicate a student’s achievement by giving a grade that is meaningful to a variety of stakeholders: students, their parents, schools, districts, and policy-makers. It is within this context of PBL curriculum that five APES teachers in this grading study were examined as they determined grades at the end of a grading period.

Innovations like this do not exist in a vacuum. Districts are complex contexts with many shifting characteristics that can interact with curricular and pedagogical changes. Of particular interest in this study was the district’s concurrent move away from traditional grading scales to implementing Standards-referenced Grading (SRG) – something that all five of the participant teachers were not familiar with and were beginning to learn about when data collection occurred. Although the district was adding a grade level a year to the SRG system, two high schools – School W and School S – decided to convert to SRG for all of their classes. As a result, the five

high schools used different grading scales despite using the same curriculum and assessments.

Studying grading practices has been complicated by differences in curricula and instructional practices that make it difficult to tease out the likely influence of school and district context. This study takes advantage of the fact that the teachers were all teaching the same curriculum, with the same or similar assignments, in the same sequence. The sample of five different school settings provides distinct contextual differences in type of student population (race, language, and socioeconomic factors), parental and community involvement, and resources available. For each of these settings, there were bound to be multiple negotiations as grades functioned as boundary objects (Nolen et al., 2011) in the different communities where grades may have different meanings, values, and uses. By including in the grading study sample five teachers from the same school district, individual differences in grading practices were observed and studied.

CHAPTER 3: Methodology

This study is a qualitative comparative case study analysis (Merriam, 2009). Participants were five teachers from the same mid-sized Midwestern school district who were teaching AP Environmental Science (APES) using the same PBL curriculum, including the same rubrics and assessments. This consistency in curriculum allowed the researcher to attend to individual teacher differences in teacher grading practices despite the uniformity of curriculum and major assessments. Their school district has a free and reduced lunch rate of 68% and was seeking to actively increase their AP course offerings as well as the diversity of their AP student population.

I have divided this section into three parts: (1) Participants and Setting, (2) Data Collection, and (3) Researcher Positionality.

Participants and Setting

As stated before, the five teachers in this grading study were selected from a larger study of project-based learning in AP classes. Each teacher taught at a different school but shared the same AP Environmental Science curriculum and attended professional development with the other APES teachers together at least once a month. In addition, because each of these teachers participated in a larger parent project about the effectiveness of PBL in AP classrooms, they attended yearly professional developments given by the university to discuss and refine the curriculum.

Tables 1 and 2 give the demographic characteristics (ethnicity, socio-economic status along with the percentage of English Language Learners and students served by Special Education programs) for each of the five studied schools.

Table 1. Demographic Characteristics of the Schools (%) – Part 1.

| Teacher | Teacher's Race | Teacher's Experience | School | Free & Reduced Lunch Rate | English Language Learners | Special Education Services |
|----------------|-----------------------|---|---------------|--------------------------------------|----------------------------------|-----------------------------------|
| Mr. Thomas | White | > 20 yrs. exp. -military background | School T | 70.7% | 7.57% | 15.23% |
| Ms. Wall | White | -3 yrs. exp. -1 st year teaching in urban setting | School W | 75.33% | 25.33% | 15.51% |
| Ms. Smith | White | -first year in classroom | School S | 84.52% | 14.37% | 23.54% |
| Ms. Nelson | White | > 10 yrs. exp. -raised in educated in Europe | School N | 52.57% | 8.72% | 12.12% |
| Ms. Martin | White | - 9 yrs. exp. | School M | 75.34% | 16.33% | 17.11% |

Table 2. Demographic Characteristics of the Schools (%) – Part 2.

| School | White | Black | Asian/ Filipino/ Pacific Islander | Hispanic | Other |
|---------------|--------------|--------------|--|-----------------|--------------|
| School T | 57% | 12% | 5% | 20% | 6% |
| School W | 36% | 28% | 15% | 16% | 5% |
| School S | 34% | 24% | 9% | 28% | 5% |
| School N | 55% | 23% | 4% | 12% | 6% |
| School M | 41% | 15% | 9% | 29% | 6% |

For each of these settings, there were bound to be multiple negotiations as grades functioned as boundary objects (Nolen et al., 2011), but because of the use of the same

curriculum that comes with the same scope and sequencing, instructional materials and tools, and common professional development sessions, individual differences in teacher decision-making when assigning grades could be investigated.

Data Collection

To understand teachers' grading practices and decision-making processes, I conducted semi-structured interviews and a think-aloud protocol with each teacher. Interviews occurred on a day the teacher participants were calculating first-semester grades. The interviews focused on specific practices teachers used to determine summative grades. For example, I asked if they adjusted some of their beginning of the year plans for instruction and assessment (and if so, how and why)? Did they feel that their grades were an adequate representation of the learning that was done daily in the their classrooms? How did they negotiate the demands of the district and others while trying fairly to assess a student on his or achievement in their class? Interview questions are shown in Appendix 1.

After an introductory interview, teachers were asked to select one high achieving student, one medium achieving student, and one low achieving student for the think aloud activity. Teachers were free to define those categories of high, medium, and low in any way. Think-aloud protocols (Ericsson & Simon, 1980) allow participants to state what they are thinking, doing, and feeling as they go about a task without filtering their thinking (which is more likely during an interview or survey) and are a valuable and “thoroughly reliable source of information about thought processes” (p. 247). Olson, Duffy, & Mack (1984) stated that think-alouds are also useful to study individual differences in performing the same task. Think-alouds were utilized as they may more specifically or veridically reveal the decision-making processes that teachers go through when trying to sum up a student's learning experience in the course with one final grade.

During think-alouds, I only used the prompt, “Please, keep talking” (Katalin, 2000) so as not to influence or structure the participants’ thoughts. As they thought-aloud, I took brief informal notes on their behavior and tone of voice. Immediately after the think-aloud, I gave a brief “exit interview” (Charters, 2003) so that the participant could help interpret the think-aloud data. In recalling the task immediately, trace memories of grading processes would be available to participants, increasing the reliability of the retrospective data (Gibson, 1997). Data, consisting of video and audio recordings of the entire grading sessions and interviews, were transcribed verbatim.

Researcher Positionality

As an “insider” to the larger study, my previous experiences with this group of teachers did affect my construction of meaning of the data in this grading study (Ganga and Scott, 2006). In this section, I will explain my interest, role, and relationship to the larger study and the participants.

Prior to beginning this grading study, I had worked with the five participants in the larger parent project where I was a research assistant. For the parent study, I filmed classroom activity in the three participating districts, coded data with the research team, and attended professional development sessions with the teacher participants. In addition, I also worked as a teaching assistant, teaching associate, and coach in the teacher education program of my university. Prior to that, I had been a classroom teacher in public schools for seventeen years and had worked as a private teacher/tutor for an additional three years.

As I collected data for the parent study, I engaged with the teachers about delivery of lessons and how lessons went. Like many reflective teachers, the participants of this study wanted to do the very best work they could so that their students would learn, succeed, and seek

further opportunities to learn. It was not uncommon for the teacher to debrief the lesson with me as I packed up my film equipment. Although I never judged a lesson to be “good” or “bad,” I did listen as the teacher reflected on the lesson and thought about next steps of instructions. Since the teacher participants knew I had been a school teacher for seventeen years, I did offer suggestions – when asked – on how to engage certain students. Despite having personal opinions about how a lesson could be improved, I never made suggestions or judged the teacher to be incompetent – instead I noted their reflectiveness even when they were highly stressed. In fact, my own long teaching career made me empathetic to my subjects as they juggled the demands of everyday routines and ambitious pedagogy in the form of PBL while being studied for a university research project. Prior to this study, I had no individual conversations with the participants about grading and summative assessment.

This previous interaction with the five participants while I was a research assistant in the larger parent project, made recruiting them for my grading study easier. All five were willing to participate in my individual study and provided me access to their gradebooks and their thoughts on grading, AP coursework, college preparation, and how best to encourage their students to pursue learning opportunities.

Analysis

Analysis of transcripts began with open coding of the interviews and the think-alouds in order to look at phenomena that may have been occurring within the data that I may not have expected. While open coding, I kept analytic memos so that I could keep track of emerging patterns in the data. As I coded, I developed categories around the existing concepts and theories that frame this study. Coding categories included themes regarding factors that influence teacher decision making while grading, constraints and affordances of grading policies, productive

friction, the negotiation of grades as boundary objects within various communities of practices, differences in grading in PBL and/or AP settings versus other settings, and other salient findings that may have emerged. When no new coding themes presented, I finalized my codes. Again, the use of the same curriculum by all five teachers allowed for findings of individual differences in grading that were unrelated to course content.

After coding, I then developed claims about what patterns I observed in the data. I considered: Do teachers consider the consequences or future negotiations that their grades will bring about in the various communities of practice where their grades serve as boundary objects? And, do they experience “productive friction” in the actual process of grading? The think-aloud procedure with each teacher helped me triangulate the interview data, along with the video data in order to determine if their self-reported plans and procedures for grading were indeed enacted in the actual moments of grading.

Following are my findings, focusing on how, in the moment, the teachers in this study produced final first semester grades and what they considered as they grappled with providing three of their students with semester grades. Findings demonstrate structure and thoughtful decision making as teachers weighed how to broker the world of college rigor and the need to keep their students engaged as learners persistent enough to move on to higher education opportunities.

CHAPTER 4: Findings

Introduction

When teachers sit down to grade, they have already done a lot of assessment work prior to grading. They assess continuously as they instruct by asking questions and checking students' work. They've already added scores to the gradebook for smaller assignments and unit tests and quizzes. So sitting down to grade, they have numerical representations of their students' learning right in front of them and it should just be a matter of using some weighting and aggregation of those grades to get the final semester grades.

The teacher participants in this study often referred to rubrics being the guide for their expectations for the assignment and what they used as a tool when grading. As an instructional practice, I observed how they each referred to the rubric of a project when discussing the meaning of how to receive a certain grade on an assignment reifying the meanings and values of grades with students, parents, and, sometimes, their school community (building level).

However, as research shows, teachers, despite warnings and mandates to grade for achievement only, self-report that they factor other considerations into a student's final grade ((Brookhart, 2011; Jongsma, 1991; Pope, Green, Johnson, & Mitchell, 2009; Randall & Engelhard, 2010). During my think-aloud sessions with the teacher participants, all teachers began with student records in their gradebook, but adjusted grades for other factors. Think-aloud and interview data revealed patterns in these judgments that give insight into their motivations.

Summary of Findings

Three main claims are made here:

1. AP teachers, because they are teaching a college level course, acted as “brokers”² for their students to help move them from the world of high school to the world of college.
2. AP teachers experienced tension between acting as brokers between the worlds of high school and college while supporting student motivation and persistence in opportunities to learn.
3. AP teachers had to negotiate the tensions of acting as brokers between the worlds of high school and college while at the same time having to adjust their grading practices in order to adhere to the district’s new policy of Standards-Referenced Grading.

Claim 1: AP teachers acted as brokers to help move their students from the world of high school to college so that their students were prepared for college.

The interviews and think-alouds provided a look into each teacher’s “black box” (Black & Wiliam, 1998) where they individually balanced and negotiated decisions in order to fairly grade their students on APES content as well as prepare their students for the journey from high school to college. The APES course content and grades served as boundary artifacts between high school expectations and the new world of college which many of the students aspired to attend. In this way, teachers brokered this journey for their students by giving grades that teachers felt aligned with AP and college expectations. Because this study took place in poverty impacted schools, many students would be first generation college students and teachers took their role very seriously in providing a vision of what being a college student entailed and the opportunity to experience higher education. The grades the teachers assigned were those boundary objects that let the student know if he or she was college ready.

² Engeström et al. (1995) describe brokers as those that “face the challenge of negotiating and combining ingredients from different contexts...” (p. 319).

Mr. Thomas felt a responsibility for students to be able to do AP work and, therefore, be ready for college. Retakes, retests, time extensions were not, in his experience and opinion, part of the college experience and he felt a duty to make sure his students were adequately prepared for college and, therefore, Mr. Thomas said he did not permit students to refine and revise their work:

Because kids need to be able to demonstrate mastery but in most colleges, they get one shot at it. I don't know that the re-testing to show mastery on the standard is appropriate in an AP course (Mr. Thomas, Interview, 15-01-14).

Mr. Thomas also used his experiences as a college student to broker the world of college for his students:

My memory of how we were taught was we were given deadlines, we were given criteria and a rubric. This is how the professor's going to grade this. We were expected to present material when we were supposed to present it. There wasn't a second chance, there wasn't a third chance. [The professor] wanted to move on (Mr. Thomas, Interview, 15-01-14).

Ms. Nelson, placed at the most well-resourced school in the district, believed very strongly that she had to prepare her students for college. When asked if she allowed students to do redos and retakes of assignments and tests:

I've spoken to professors at [the state university]...and they don't allow them. I just don't want it to be a huge shock when they become freshmen (Ms. Nelson, Interview 15-01-13).

Ms. Nelson wanted her students' grades to prepare them for college; she even inquired into other academic communities of practice so she could make sure her grades as boundary objects were meaningful and valued at the college level. It was important to her that her students be college ready and that she stick consistently to the AP rubrics so that she would be grading her students consistent with the idea that the AP College Board did indeed prepare students for success in college. She also felt a responsibility to her students, families, and school community for getting her students into college (Ms. N., interview, 15-01-13).

Claim 2: AP teachers experienced tension as they acted as brokers between the worlds of high school and college while supporting student motivation and persistence in opportunities to learn.

However, not all teachers in this study disallowed redos and retakes although they all expressed a desire to make sure that their students were college ready. These teachers balanced the responsibility to get their students ready for college, but also balanced the need to stay connected to their students so that they would learn the content and persist as learners. Figure 3 shows a representation of the structure to some of the participants' decision-making as they determined first semester final grades. It is, in some ways, indicative of Bishop's (1992) notion of the difficulty of putting the teacher in the "judge" (far left of the scale) and "coach" (right side of the scale) roles simultaneously. This representation displays the weight on these teachers to play the "judge" and represent what they envision college will entail for their students with the need to advocate ("coach") their students towards future learning opportunities – to mentor them to persist and see that they too can learn.

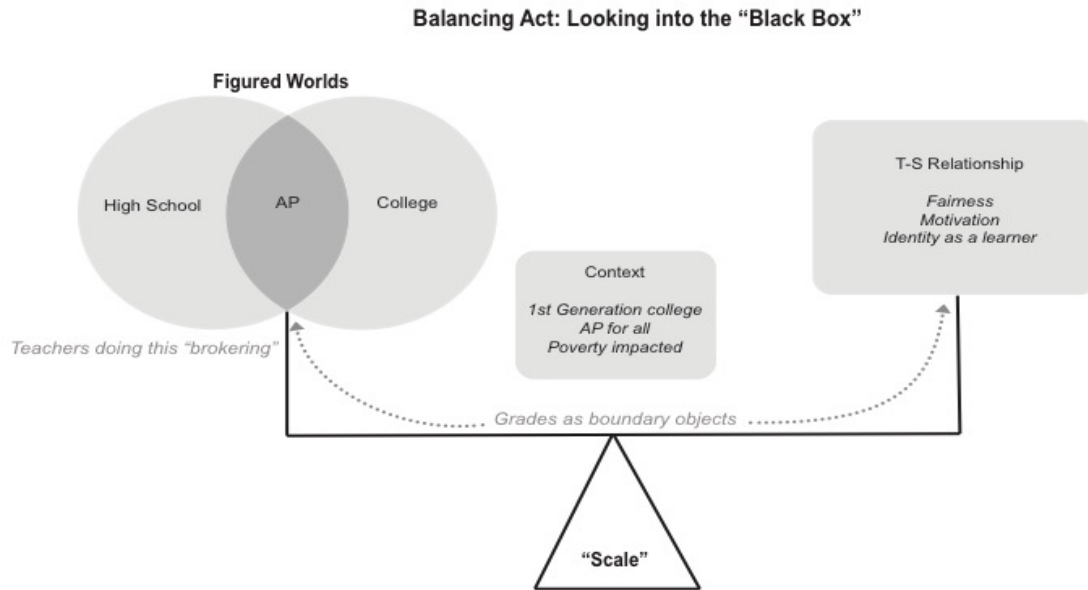


Figure 3: Summative grading as a “balancing act.”

“Boosting” Grades

Ms. Smith, working in the most poverty-impacted school, which was piloting the SRG policy, worried if she was truly representing the world of college for them. She wanted to be a tough grader but worried about the effect on her students’ motivation. This was a tension, she confessed that she daily worried about:

I don’t want to grade easily and give them all good grades, because that will help them right now, but in the future? “Yeah, I got a B in my AP class.” Then they go to college and totally bomb it. The same way, I have these really high expectations because I took a lot of environmental science classes in college and this is the stuff that I had to do and this is a college level class and they’re not doing that... Then does that A mean they really should have gotten an A? I don’t know. Those are the things that weigh on me. The

biggest (Ms. Smith, interview, 15-01-13).

Ms. Smith felt that the onus was on the teacher for the students' learning and allowed retakes of assignments so she could reteach and/or give her students more time for practice:

Sometimes I feel like I haven't taught them well enough for them to do well, and do I hold that against them? I can't. That's on me. Then does that mean they really should have gotten an A? I don't know. Those are the things that weigh on me... The only policy we have here is retakes and re-do's and that's a school wide policy that we have written in every department (Ms. Smith, Interview, 15-01-13).

Ms. Smith felt a responsibility to lead her students – many of whom would be first generation college students - to post high school learning opportunities. However, she always balanced her ability to communicate concepts and her students' level of learning as equally important. Grades were boundary objects between her and her students that communicated how well her students learned and how well she facilitated the students' learning.

During the semester, when earned grades prompted concerns about discouraging students, a teacher might act to manipulate grades by adding or modifying assessments before the end of the semester. For example, Ms. Smith worried about the effect students' grades would have on their motivation to continue on in the AP course. For this reason, she told me, she would give an easy vocabulary matching test as the final exam (something not part of the APES assessment curriculum) in order to boost the students' grades:

I grade the FRQ (Free Response Questions) pretty close to ...like in the AP, how they grade it. I tried to grade it pretty close to that.

Some feel like it's a little bit too hard. I'm not a college level teacher here. I feel your thinking, I like your thinking, I'm going to give you points for your thinking...For the final a boost is better. Motivate them for the second semester (Ms. Smith, Interview, 15-01-13).

Ms. Nelson, who did not allow retakes on assignments and tests, stated that she assessed her students on achievement only. Her way of manipulating grades, when she received push-back from parents or the school community, was to allow extra credit to boost grades.

Ms. Nelson, placed at the school with the lowest free and reduced lunch rate, used traditional grading (in fact, all of her assessments were graded on a curve), but at the same time felt there was a lot of pressure "to improve the graduation rate" and to be held accountable to parents. She was very aware what her grades communicated to her students, her school, parents, and the district:

I think individual buildings want students to be successful, and if students are failing I feel like people look at what did I not do to help the student be successful? What more should I have done, or could? When I grade and look and there's a class of twenty-four, six or seven Fs, yeah. I've heard teacher stories of teachers being called into the office and asked those questions. It's not personally happened to me but I'm aware that it might (Ms. Nelson, Interview, 15-01-13).

Ms. Nelson stated that her policy was not to allow retakes of assignments and tests. But as the end of the semester neared, she stated that she traditionally received more phone calls

from parents regarding their children's grades:

At the end of the day, I really believe I give students all the opportunities in the world, but if they consistently don't come to class, if they consistently don't hand things in, at the end of the day, they need to figure out that they have to do something in order to get something at the other end (Ms. Nelson, Interview, 15-01-13).

As stated earlier when comparing Ms. Smith's and Ms. Nelson's manipulation of grades, Ms. Nelson "managed" the tensions between parents and students was by allowing extra credit at the end of the semester when grades were not considered high enough by students and families. Ms. Nelson did not define what an "acceptable" grade was, but in her school community, the emphasis by parents, teachers, students, and administrators to have students attend college meant that grades were much more than indicators of learning; grades also served a "gate-keeping" role – grades translated who was eligible for college learning – something very important in Ms. Nelson's community.

Playing the "Judge" vs Being the "Coach"

Teachers often see themselves in the dual roles of "judge" and "coach" when they grade (Bishop, 1992). They have to evaluate a student's progress and judge if a student is proficient in some area of learning. They also have to advocate for their students' learning by "coaching" them along and providing incentives to persist in learning. In his interview, Mr. Thomas saw himself as a broker between the worlds of high school and college as he prepared his students for college. But during his think-aloud, he also talked about his role in the classroom as if he were more of a "coach" or mentor of young people still on the path to adulthood where poor decisions

are part of the journey. In the end, Mr. Thomas accommodated his AP students' needs, even though in his introductory interview he claimed to not support retakes and retests for his students wanting to go on to college.

Students will come in and they will, last week of the semester or grading period, want to turn in a lot of work and I'm very willing to let them do that, but I also caution them on their method and their approach to success. I work with the student because I figure, these are still kids. They're still kids and they're still liable to make poor decisions. You have to work with them (Mr. Thomas, Think-Aloud, 15-01-14).

Ms. Nelson was dedicated to making sure her class was "college-level" so she consistently used AP rubrics as she graded and stressed the importance of the rubrics to her class so they could pass the AP exam and, therefore, receive college credit. [Note: In all of the interviews, Ms. Nelson was the only teacher to discuss her desire for her students to perform well on the AP exam. The other teachers expressed a desire for their students to do well in class and to go on and do well in college.]

Ms. Nelson, when grading, very much took of the role of "judge"/evaluator of the students' mastery of the material. Throughout Ms. Nelson's think-aloud while grading final projects for the first semester, she used an AP examiner's lens as she assessed student work. She repeated, "Would the AP examiner know what [he/she] means?" (Ms. Nelson, Think-Aloud, 15-01-13) and used that as the basis for how many points she would assign. Ms. Nelson used the PBL curriculum and College Board tools to align her assessment methods with the College Board's AP expectations so that her grading would be a reflection of who deserved college credit

and who will go on to do well in college. But because Ms. Nelson's different communities of practice, in this case, the family and the school community, had placed very high value on grades, Ms. Nelson offered extra credit so that students could bring their scores up. So although Ms. Nelson remained consistently on the left side of the scale in Figure 3 (brokering the worlds between high school and college), she still did consider other factors than achievement as a compromise while working at the boundaries of the family's and school community's worlds.

Use of Informal, Anecdotal Assessment Data While Formally Assessing Students

Teachers often use other informal, sometimes more anecdotal information when summatively assessing. In addition to the gradebook, Ms. Martin considered other, informal sources of information about student learning when grading her students. During her think-aloud, when Ms. Martin saw a student struggling to explain their learning on their written assignments, Ms. Martin would often think back to learning episodes in the classroom to assess her students' actual learning:

Just by her mentioning the decreasing property values, yeah, that was something we did mention in class. That wasn't anything really cited. It was something we talked about in class. She understands the relationship between the two (Ms. Martin, Think-Aloud, 15-01-14).

Grades can have great motivational meaning by providing successes to students about their ability to learn (Briscoe, 1994; Brookhart, 1994; Campbell, 2012; Harrison, Konings, Schuwirth, Wass, & Vleuten, 2014; Thomas & Oldfather, 1997). Ms. Wall thought it important that her grades give students a chance to see themselves as smart and capable of doing AP work. During the grading think-aloud, Ms. Wall, who had many students who were English Language

Learners in her APES class, thought back to times in class when connections were made during class discussions and used those moments to boost the student's grade. When asked how she used "professional judgment" to decide on a final semester grade, she responded:

I can't be like, "This kid's such a butthead. I'm going to give him a 2." When I'm thinking about it, it's only like she randomly came up with the fact, "Hey, when I lived in Liberia, our water usage were so much lower." I'm like, "Okay, you're adding on another factor to the whole water standard. I know you are making these connections and really thinking about things." If I'm debating between a 2 and a half and a 3, then I'm like, okay probably a 3, because she's making connections. She's had experiences (Ms. Wall, Think-Aloud, 15-01-14).

This quote from Ms. Wall is very similar to what Ms. Martin did consistently as she graded her students from the first semester. During her grading think-aloud, Ms. Martin referred back to actual classroom experiences for each of the students she graded as I was observing. She felt this a fairer and more accurate way to assess since "the assignments in AP are not very quantitative...[and] there's always shades of gray with the AP course" (Ms. Martin, Think-Aloud, 15-01-14). She further explains why she thinks back to classroom experiences as she grades:

I have noticed that just even during class when we're having discussions, when we're doing the reading and we're talking about it, just the group discussions or class discussions that we hold, really lets me know if they're actually following along or where they missed a certain understanding of a particular topic (Ms.

Martin, Think-Aloud, 15-01-14).

For Ms. Martin, assessment in the moment served to provide a fairer grade that cannot be quantified with PBL activities, and it also helped inform her instruction in the case of misconceptions and mistakes regarding the content. Ms. Martin's grades served formative as well as summative functions for her.

Claim 3: AP teachers felt tensions between their roles as brokers between high school and college and the grading practices mandated in the district's new policy of Standards-Referenced Grading (SRG).

Each participating teacher was dealing with the change mandated by the district to move from a traditional grading system to Standards-Referenced Grading system – something that some of the participants felt they had not been adequately prepared for. This is not unlike the tension felt by teachers in other school districts caused by changes in curricula or instructional practices or, perhaps, newly mandated policies that often occur during a school year. Just as the teacher participants were getting used to teaching open access AP in their neighborhood schools, without a testing or vetting procedure to allow students access to the AP class, they struggled with the demands that soon they would have to, or in two cases, already had to, change how they would summatively assess students.

Three of the teacher participants (Mr. Thomas, Ms. Nelson, and Ms. Wall) expressed doubt about how they would follow the tenets of SRG while maintaining the integrity of an Advanced Placement course:

I don't know [how I will handle that] yet, because one of the tenets of standard-referenced grading is that kids can retest to show mastery on a standard. Unless the colleges are doing that, which I

don't think they are, I don't know how applicable [standards-referenced grading] is to an AP course (Mr. Thomas, Interview, 15-01-14).

Ms. Nelson was unsure of how she would handle SRG when it came to her AP classes. She did not believe it was good practice for college bound students and firmly believed in not allowing her students retakes and redos when they struggled (Ms. Nelson, Interview, 15-01-13).

Ms. Wall, placed at a school where they were already piloting SRG, questioned the appropriateness of open access to AP:

Things that weigh on me is [the district's] all inclusive for the AP classes. They technically don't have any prerequisites. [Biology is the only prerequisite for APES.] You can be taking a 5th grade elementary level math and still be okay to join the AP environmental class...When I'm grading should I be grading them on their math skills? Should I give them the formulas to do these? Or should I [have them] tough it out on their own because that's not what they're going to get on the AP test? (Ms. Wall, Interview, 15-01-14).

Although Standards-Referenced grading is a system designed to grade students' achievement and is different from the familiar traditional 100 point grading scale,, there are many positive aspects and outcomes from using an SRG system, many of which enhance student motivation and self-knowledge as a learner. Some examples:

- Students can retake an assignment or test at any time. Showing they have learned the concepts and content are the main goals.

- When students do not complete an assignment, a zero is not given so that it does not plummet a student's grade when it is averaged in with other scores; instead, the student is encouraged to complete the assignment or activity when they are ready.
- Students frequently self-assess their progress which increases intrinsic motivation and self-regulation of learning. (Author's personal notes from this district's professional development session on moving forward with Standards-Referenced Grading.)

Standards-Referenced Grading can be a very trustworthy, honest way of communicating with the student about how well they are learning. Some of the tenets of SRG can provide the motivational balance to increase student self-confidence as a learner so that a teacher can more easily prepare that student for the rigors of a college education.

Summary of Findings

All five of these teachers graded their students quite differently despite the same curriculum, assessments, and professional development meetings. But, as seen in Figure 1, there was a structure to their decision making as all (except Ms. Nelson) alternated back and forth about using grades to confirm if they were ready for the world of college (as interpreted through the teacher's experiences) and using grades so students would persist as learners and seek further opportunities to learn. Ms. Nelson, teaching in a more well-resourced school with more parent support and expectations, perhaps did not have to worry about motivating her students to learn and focused on the job of brokering the world of college and preparing her students for a successful college career. Contrast this with Ms. Smith, teaching in the most poverty-impacted school where she gave her students a "boost" on the final (a matching vocabulary test) so that her students would not become discouraged and drop the course after the first semester and, thereby, never pursue future opportunities in higher education. When Ms. Nelson gave students a "boost"

to their grades, it was in the form of extra credit often requested by families and students dissatisfied with their grades.

Although most of the teachers, at the beginning of each interview, said they graded on achievement scores (regardless if they used a traditional or SRG grading method), each teacher went on to show me, as they thought aloud during grading, the other elements they considered when giving a final summative grade.

Although all of the participants of the study had different ways of individually grading students and had their gradebooks set up differently, four of the teachers all had similar structure to their thought processes when deciding on a grade for the semester. Four of the five teachers balanced the need to act as broker and represent the world of college using grades to communicate their readiness with the need to motivate their students to continue and persist in opportunities to learn. The fifth teacher also used grades to communicate if her students were ready for college, but perhaps because she was employed in the most well-resourced school, there was no need for her to use grades as way to motivate students to higher education opportunities.

The teachers in this study grappled with the following while grading: 1) They acted as a broker for their students so their students would understand the demands of college, 2) they implemented PBL practices in a class traditionally taught through lectures and textbook readings and had to design the room and motivate the students for the PBL activities, 3) they had to monitor group work in order to assess each student's progress, 4) they worked for a poverty-impacted district encouraging their students to enroll in an AP class so a greater majority of their students would have AP opportunities, and 5) they were in the midst of the district changing the grading policy to Standards-Referenced Grading. Throughout this, grades continued to serve as

boundary objects with multiple purposes depending on the community of practice affected by the grades.

CHAPTER 5: Discussion and Conclusion

In this final chapter of the dissertation, I will discuss findings from Chapter 4, summarize their implications for assessment research, present limitations of this dissertation study, and suggest future directions for research on summative grading. This chapter is divided into two sections: discussion and conclusion. The discussion section includes three subsections, each dealing with a claim in my findings. The second major section of this chapter is the conclusion section and is divided into two parts. The first part presents implications and future research directions the limitations of the current study, while the second part provides the limitations of the current study.

Discussion

The teachers in this study originally participated in the parent study because they wanted to try out PBL practices, which stress conceptual and deeper learning in the context of an AP class where traditional instruction is still the norm. Along with adopting PBL instruction, the five teachers in this study had to contend with the AP world: large amounts of knowledge to be learned and the AP exam, which provides possible future college credit. In addition, many of the students they were teaching would be the first generation of their families to attend college. By expressing their thoughts about grading – as they were grading – the teachers revealed the tensions and negotiations they experienced as well as the structure of their decision making (See Appendix 2).

Claim 1: AP teachers acted as brokers to help move their students from the world of high school to college so that their students were prepared for college. Grades, for the teachers in this study, functioned as boundary objects (Nolen et al., 2011) giving students in high school information about the world of college. Given that this study takes place in the context of

an AP course, grades were indicators of how well the students – at the high school level – were achieving in a simulated college course. The meanings of these grades were important and had value in communities of practice where students attending college has particular meaning and value. The teachers in this study acted as “brokers” (Akkerman & Bakker, 2011) – mediating for students the difference between the worlds of high school and college. Nolen et al., (2011) described grades as artifacts as “tools for learning and development” (p. 89). In this case, grades were the tools for learning about how college works and what colleges expect. Engeström et al. (1995) described brokers as those that “face the challenge of negotiating and combining ingredients from different contexts...” (p. 319). One of those contexts was the teachers’ own experiences as college students. For example, Mr. Thomas remembered that students only got one shot at tests and assignments and retakes were not allowed when he went to college. He felt that he would not be preparing his students for college or giving them a realistic view of how college operates if he did not broker the world of high school to college using his own experiences. The same is similar for Ms. Nelson who consistently graded her AP students on a curve. As a college student, she was graded on the curve (Ms. Nelson, Interview, 15-01-13) and she would be remiss if she did not prepare her students for this kind of assessment in college.

For all five of the teacher participants, acting as brokers between the world of high school and the world of college was a part of their job because their AP Environmental Science course was considered college level and their students would be eligible to obtain college credit at the end of the course if they passed the AP Environmental Science exam. The think-aloud data provided evidence of the importance of this part of their job, as well as how each teacher’s individual memory and/or vision of what the college experience entailed influenced their grading decisions.

The design of this study using a think-aloud protocol allowed a glimpse into the structure of teacher decision-making as teachers struggled with brokering the world of college with providing feedback to their students that they can learn and that they are students who can continue on into higher education.

Claim 2: AP teachers experienced tension as they acted as brokers between the worlds of high school and college while supporting student motivation and persistence in opportunities to learn. As much as the teachers represented the world of college for their students, four of them spoke of the tension inherent in preparing their students for the world of college while keeping their students motivated to learn college-level material. Project-Based Learning is intended to be engaging, with its inquiry-based structure and use of collaborative learning and authentic projects. However, the classes studied were still Advanced Placement with a breadth of knowledge and concepts to be covered before a high stakes AP exam. The teacher participants worried about students being discouraged by poor grades when they were doing such challenging work, and teachers often offered a “boost” (i.e., Ms. Smith’s vocabulary final, Ms. Nelson’s extra credit, and other teachers allowing retakes on assignments and tests) so their students’ semester grades would keep them motivated to keep learning. This is also true of Ms. Martin when she would think back to class learning experiences and provide anecdotal data while grading so she could look at the student more “holistically” – not just relying on the numbers in the gradebook. By remembering moments of learning in the classroom, Ms. Martin gave her students a “boost” while grading as she reflected on those past moments.

Cheng and Sun (2015) point out the duality of the teacher’s role when grading: they discuss teachers having to be in both the role of “judge” and “coach” – two very difficult roles to fulfill when grades for students are used in so many capacities (i.e., entry to certain courses of

studies, access to college, family approval, providing a student with a sense of self-efficacy, and even lowering a student's car insurance rate!). Bishop (1992) argued that putting the teacher in the place of both judge and coach is unfair to teachers, given their many responsibilities, and leaves students without an advocate for their learning needs. For many students in the teacher participants' classrooms, their socio-economic status or status as a refugee or immigrant perhaps necessitated an advocate in the classroom for their learning needs. It was perhaps because Ms. Nelson taught in a more well-resourced school compared to the other four schools that she did not mention grades as way to motivate her students to continue learning.

Claim 3: AP teachers felt tensions between their roles as brokers between high school and college and the grading practices mandated in the district's new policy of Standards-Referenced Grading (SRG). When grading, teachers must also consider practices and policies mandated by their school communities while at the same time managing all the planning, instructing, and assessing they do. The district in this study was not only in the process of expanding access to AP for their diverse student population, they were also piloting a new grading system. The teachers in this study - like most teachers, in general - already had so much to do: working with a Project-Based Learning curriculum in a traditionally taught class, getting their students ready for the AP exam, preparing them for college – the addition of a new grading system seemed overwhelming for some of the teachers in this study.

Research suggests that the dilemma faced by these teacher participants is not unique. Black and Wiliam (1998) ask, "Are we serious about raising standards?" (p. 90) and again, I repeat their concern:

Teachers will not take up ideas that sound attractive, no matter how extensive the research base, if the ideas are presented as

general principles that leave the task of translating them into everyday practice entirely up to the teachers (p. 87).

Considering the input from different communities of practice is valuable, argue Black and Wiliam (p. 88), but it is not until there is support for teachers at the classroom level that assessment will begin to be meaningful for all stakeholders:

Policy ought to start with a recognition that the prime locus for raising standards is the classroom, so that the overarching priority has to be the promotion and support of change within the classroom (p. 88).

Until there is an understanding of the conditions in which teachers do their work and how different communities of practice can create additional tensions for teachers, learners' needs will not be supported. Assessment will continue to function at a superficial or gate-keeping level instead of becoming important feedback for students to improve learning, helping them to set goals for learning, and experience the self-efficacy and identity of being a learner.

Conclusion

This dissertation investigated how teachers navigated and negotiated tensions between communicating expectations for college level work and students motivation as they determined grades in an AP level Environmental Science course.

Cheng and Sun (2015) state that

It is essential to take into account teachers' grading decision making...because the consequences of grading on students'

learning are teachers' primary concerns, [and] the factors they consider in making grading decisions reflect their beliefs and assumptions about learning, which are in turn influenced by...social, cultural, and educational contexts (p. 231).

Limitations. There are, of course, limitations to this study. A larger number of participants were sacrificed for thicker, multi-sourced data on teachers' enacted grading practices. Although the context of this study is a very specific one (same curriculum and project based learning), that very specificity provides control over some variables that might make it difficult to draw any conclusions on the primary variable of interest - teacher grading practices. So although there are inherent limitations in a small sample, there are also advantages in looking at the commonalities and variations between a group of teachers delivering the same curriculum and situated within the same district and how they think about grades. Observing only semester grading might have affected the results. Grading mid-year might have influenced teachers' thinking. It would be valuable to do the same think aloud at the end of the year to see if teachers changed their positions or strategies once the final grades were to be assigned.

Implications. Investigating the "black box" - full of input from various communities of practice and then looking at what teachers are actually considering when they grade can give insight into how teachers are engaged in a balancing act. As they strive to facilitate rigorous learning opportunities for their students, they also have to motivate their students to persist in challenging learning opportunities, while at the same time, communicate what the future world of college expects. Investigating this tension in different communities of practice, discussing the way their grades serve as boundary objects in a variety of ways, could possibly provide the "productive friction" (Ward, Nolen, & Horn, 2011) needed to "initiate positive changes in [the]

use of high-leverage practices to improve student learning and understanding” (p. 15) and provide consensus of what it means to be a successful AP Environmental Science student. Black & Wiliam (1998) suggest that given the increasing inputs from other communities of practice, teachers need support and understanding from other communities of practice. They suggest that policy makers and other communities of practice need to be part of the discussion on how best to support teachers and learners in the area of summative assessment.

There are implications for teacher education and professional development programs as well. Despite the recommendations in the literature to grade for achievement only, Brookhart’s 1993 study showed that teachers still grade for other factors - even when they receive assessment and measurement training. Instead of telling teachers to grade for achievement only and here is how to do it, perhaps, it is best we listen to teachers and hear why they make the decisions they make when grading. Assessment courses and district professional development classes can benefit by expanding the meaning of assessment so that it includes consideration of other factors besides achievement or mastery of given concepts and skills. Perhaps then teachers will feel more comfortable and more transparent about their role as a “coach” or an advocate for their students’ needs when grading. There is no shortage of standardized, external assessments for students. Why not allow those tests to play the “judge” role in assessment (Cheng and Sun, 2015)? External assessments are only one small piece of the picture when looking at a student’s progress. Teaching teachers and preservice teachers how to assess as an advocate for their students, as opposed to acting as a judge of learning, can perhaps make feedback in assessment more meaningful for students so that students will be more likely to take up feedback for future learning opportunities.

Part of that move toward “productive friction” could include a conversation about what

the purpose of grades are in each community of practice. Is it to give the message that a student has a certain level of mastery of, in this case, the concepts of college Environmental Science? Or do grades reveal how well a student can manage large amounts of college level material and successfully complete a college level exam? Or do we give grades because they are needed for students to attend college? This study, by using an interview and think-aloud protocol, provided a look into how these five particular teachers constructed the meaning and value of grades. By considering the various constructs different communities of practice have of grades, we can then look at the criteria teachers use and give to their students and families with the hope of making grades' meanings more accessible to different communities. Brookhart et al. (2016) state that when criteria for grades are understood by students and others, grade meaning is "enhanced" and "the dimensions on which grades are based should be defensible goals of school and should match students' opportunities to learn" (p. 836).

Thinking of grades as boundary objects among different invested partners in learning allows all to hopefully see assessment practices and tools as "potential bridges to communication but also as tools for learning and development" (Nolen et al., 2008, p. 89) for students, teachers, and all who hope to encourage more equitable access to AP courses. Teachers have real reasons for why they assign the grades they do. Sharing their decision making with other communities of practice who also use grades is imperative for forward progress toward trustworthy and motivating assessment practices. Productive friction at these different groups' boundaries can lead to the discussion that could possibly align each group's expectation of the goals, values, and meaning of grades so that future AP students find learning the content meaningful, persist in other opportunities to learn, and are prepared for the world of college.

References

- Akkerman, S.F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*, 132-169.
- Allen, D., Ort, S.W., & Schmidt, J. (2009). Supporting classroom assessment practice: Lessons from a small high school. *Theory Into Practice, 48*(1), 72-80.
- Bass, K.M., & Glaser, R. (2004). *Developing assessments to inform teaching and learning: CSE Report 628*. Washington, D.C.: US Department of Education.
- Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership, 49*(6), 15–18.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81-90.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 21*, 5-31.
- Boss, S., Johanson, C., Arnold, S.D., Parker, W.C., Nguyen, D., Mosberg, S., Nolen, S., Valencia, S., Vye, N., & Bransford, J. (2011). The quest for deeper learning and engagement in advanced high school courses. *The Foundation Review, 3*(3), Article 3.
- Briscoe, C. (1994). Making the grade: Perspectives on a teacher's assessment practices. *Midwestern Educational Researcher, 7*, 14-25.
- Brookhart, S.M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279-301.
- Brookhart, S. M. (1991). Letter: Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35–36.

- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123–142.
- Brookhart, S.M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, 10(2), 161-180.
- Brookhart, S.M. (June 2011). *The use of teacher judgment for summative assessment in the United States: Weighed in the balance and (often) found wanting*. Paper presented at the Invited Research Seminar - Teachers' Judgments with Systems of Summative Assessment: Strategies for Enhancing Consistency, Oxford University Centre for Educational Assessment.
- Brookhart, S.M., Guskey, T.R., Bowers, A.J., McMillan, J.H., Smith, J.K., and Smith, Lisa F. et al. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803-848.
- Campbell, C. (2012). Learning-centered grading practices. *Leadership*, 41(5), 30-33.
- Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education*, 12(2), 68-82.
- Cheng, L, and Sun, Y. (2015). Teachers grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12, 213-233.
- Darling-Hammond, L., Barron, B., Pearson, P.D., Schoenfeld, A.H., Stage, E.K., Zimmerman, T.D., et al. (2008). *Powerful learning: What we know about teaching for understanding*. San Francisco: Jossey-Bass.

- Engeström, Y., Engeström, R., & Kärkkäinen, M. (1997). The emerging horizontal dimension of practical intelligence: Polycontextuality and boundary crossing in complex work activities. In R. Sternberg & E. Grigorenko (Eds.), *Intelligence, heredity, and environment* (pp. 440–462). New York, NY: Cambridge University Press.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Whittrock (Ed.), *Handbook of research on teaching* (3rd ed.). Old Tappan, NJ: Macmillan.
- Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251.
- Fisher, D., & Atkinson-Grosjean, J. (2002). Brokers on the boundary: Academy- industry liaison in Canadian universities. *Higher Education*, 44, 449–467.
- Ganga, D., & Scott, S. (2006). Cultural “insiders” and the issue of positionality in qualitative migration research: Moving “across” and moving “along” researcher-participant divides. *Forum: Qualitative Social Research*, 7(3), Art. 7, 1-11.
- Gibson, B. (1997). Taking the test: Using verbal report data in looking at the processing of cloze tasks. *Edinburgh Working Papers in Applied Linguistics*, 8, 54-62.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy, & Practice*, 4(3), 365-379.
- Harrison, C.J., Konings, K.D., Schuwirth, L., Wass, V., & Vleuten, C. (2014). Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences*, June 2014.
- Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420.

- Jongsma, K.S. (1991). Rethinking grading practices. *The Reading Teacher*, 45(4), 318-320.
- Katalin, E. (2000). "Please keep talking": The 'think-aloud' method in second language reading research. Retrieved from www.deal.elte.hu/pages/novelty/htm2/vol73/elekes.htm
- Kirton, A., Hallam, S., Peffers, J., Robertson, P., & Stobart, G. (2007). Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools. *British Educational Research Journal*, 33(4), 605-627.
- Lewin, T. (2010). Expansion of A.P. tests also bring more failures [Electronic Version]. *The New York Times*. Retrieved from www.nytimes.com.
- Marzano, R.J. (1996). *A comprehensive guide to designing standards-based districts, schools, and classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mellon, K. (March 25, 2010). HISD considers mandatory policy to ensure courses are rigorous enough. *The Houston Chronicle*. Retrieved from www.chron.com.
- Merriam, S.B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- National Research Council (2002). *Learning and understanding: Improving advanced study of mathematics and science in U.S. high schools, Executive Summary*. Washington, DC: National Academy Press.
- Nicol, D.J., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Nolen, S.B., Horn, I.S., Ward, C.J., & Childers, S.A. (2011). Novice teacher learning and motivation across contexts: Assessment tools as boundary objects. *Cognition and Instruction*, 29(1), 88-122.

- O'Connor, K.B. (2009). *How to grade for learning*. Newbury Park, CA: Corwin.
- Olson, G.J., Duffy, S.A., & Mack, R.L. (1984). Thinking-out-loud as a method for studying real time comprehension processes. In D.E. Kieras & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 253-286). Hillsdale, NJ: Erlbaum.
- Parker, W., Mosberg, S., Bransford, J., Vye, N., Wilkerson, J., & Abbott, R. (2011). Rethinking advanced high school coursework: Tackling the depth/breadth tension in the AP "US Government and Politics" course. *Journal of Curriculum Studies*, 43(4), 533-559.
- Pope, N., Green, S.K., Johnson, R.L., & Mitchell, M. (2007). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education*, 25, 778-782.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26, 1372-1380.
- Riley, M. (2005). *AP as the 'common curriculum'*. Retrieved from <http://apcentral.collegeboard.com/apc/public/features/18762.html>.
- Saint John's University. (2017). Decision Maker Flow Chart. Retrieved from <https://www.csbsju.edu/education/teacher-as-reflective-decision-maker/conceptual-framework/decision-maker-flow-chart>
- Schneider, J. (2009). Privilege, equity, and the Advanced Placement program: Tug of war. *Journal of Curriculum Studies*, 41(6), 813-831.
- Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Star, S.L., & Griesemer, J.R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science (Sage)*, 19(3), 387-420.

- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22-37.
- Taylor, C.S., & Nolen, S.B. (2008). *Classroom assessment: Supporting teaching and learning in real classrooms* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: "That's my grade. That's me." *Educational Psychologist*, 32(2), 107-123.
- Ward, C.J., Nolen, S.B., & Horn, I.S. (2011). Productive friction: How conflict in student teaching creates opportunities for learning at the boundary. *International Journal of Educational Research*, 50, 14-20.
- Wenger, E. (1998). *Communities of practice: Learning, Meaning, and Identity*. New York, NY: Cambridge University Press.
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3 & 4), 283-289.

Appendix 1: Interview Questions

Pre-ThinkAloud Questions

I have a few questions for you before we begin your think aloud.

Can you give me a quick run-down on how you came to this position including your educational background?

- **How long have you been teaching APES? PBL? APES in a PBL setting?**
- **Probe: How different is this educational setting from the one you learned in?**
- **Probe: What keeps you teaching here at this school?**

What is your background in assessment/grading practices?

- **Probe: Did your teacher education program prepare you for assessment? District PDs?**

Does your school/district have specific grading policies?

- **Probe: Are there any different policies for grading AP students?**

With whom do you discuss grades and grading?

- **Probe: Coworkers, school/district committees, community?**
- **How are these conversations different?** When you give grades to your students, what factors do you consider besides numerical scores on assignments and tests? What are some of the tensions you struggle with when grading?

ThinkAloud

“Please, keep talking...”

Exit Interview

Anything you'd like to clarify about the think aloud process?

You just graded three distinctly different achieving students. Did you have a different strategy for grading each?

Do you give group grades for PBL projects? If so, how does that work?

• Probe:

- o **Accountability issues**
- o **parent or student complaints**
- o **increase or decrease in engagement – productive disciplinary engagement or in “doing school”**

Do homework assignments affect final semester grades?

In August, we interviewed you when you were in Seattle for PD and you discussed how you were going to grade your students this year. Have any of your plans for grading changed during the first semester? If so, why?

Do any of the daily PBL tasks get counted in the final grade or is it just the final project that gets summatively reported?

(From first interview): Prompt for:

- **stuff that would go in a syllabus or grading policy (how do you assign grades on progress reports or report cards?)**
- **How do you decide which things get counted in the final grade,**
- **how do you assign weights to different things (project papers, LCA [Life Cycle Analysis] posters, tests, etc.)**
- **Is that similar or different to your other (non PBL) classes? If this is the first time to**

teach using PBL, will that affect your grading practices? Thank you for your time. I may have further questions for you later in the year that I can email you.

Will that be ok?

Appendix 2: Summative grading as a “balancing act”

