Community College Full-Time Faculty Members’ Perceptions of Course Grading

April C. Zawlocki
april.zawlocki@gmail.com

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Part of the Community College Education Administration Commons, and the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
Zawlocki, April C., "Community College Full-Time Faculty Members’ Perceptions of Course Grading" (2021). Graduate Research Theses & Dissertations. 7803.
https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations/7803

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
ABSTRACT

COMMUNITY COLLEGE FULL-TIME FACULTY MEMBERS’ PERCEPTIONS OF COURSE GRADING

April Zawlocki, Ed.D.
Department of Curriculum and Instruction
Northern Illinois University, 2021
Laurie Elish-Piper, Director

This study examined community college full-time faculty members’ perceptions of grading and reporting student learning via course grades using a mixed methods approach. The research questions that guided this study asked about the differences that exist in community college full-time faculty members’ perceptions of grading by different academic subdivision, by having or not having a degree in education, and by gender. Additionally, the study examined how perceptions of the purpose of grades differ among faculty members. The study also sought to find out which criteria were used to determine course grades and what experiences faculty members attribute to learning how to grade. For the study, a sample (N=133) of full-time faculty members participated in a one-shot survey, and the data were analyzed using inferential statistics. Subsequently, semi-structured interviews of full-time faculty members (N=8) were conducted to analyze the qualitative data. From the quantitative data, a difference was found in how faculty members grade based on academic subdivision. Additionally, there was a difference found in the purpose of grades based on the number of years the faculty member has been in the field of education. In the interviews, participants reported a wide variety of experiences
regarding how they learned to grade, as well as a wide variety of components in their grades, which raised concerns regarding the consistency and accuracy of course grades.

*Keywords:* Community college, higher education, assessment, grading, scoring, achievement rating, college faculty, teacher attitudes, teacher opinions, perceptions of grading
COMMUNITY COLLEGE FULL-TIME FACULTY MEMBERS’
PERCEPTIONS OF COURSE GRADING

BY
APRIL ZAWLOCKI

© 2021 April Zawlocki

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
DOCTOR OF EDUCATION

DEPARTMENT OF CURRICULUM AND INSTRUCTION

Doctoral Director:
Laurie Elish-Piper
ACKNOWLEDGEMENTS

It is with sincere gratitude that I acknowledge the commitment of my committee, Dr. David Walker and Dr. Sally Blake, and especially my chair, Dr. Laurie Elish-Piper. Even in a pandemic, they were supportive and coached me through this process. I swear Dr. Elish-Piper does not sleep; she’s some sort of super-human. I cannot accurately describe how thankful I am for this group of individuals, without whom, I never would have completed this process.

When I started this journey, over 9 years ago, I had just found out I was pregnant with my third child, Amelia. My older two children, Trevor, was 5 at the time, and Sophia, was 2. Thankfully, I have an incredibly supportive husband who cared for the kids while I went to class after class and while I spent countless hours reading and writing. Over the years, my children grew up, and as I write this, Amelia is now 8, Sophia is 11, and Trevor is 15. It’s hard to believe all of these years have gone by. Trevor, Sophia, and Amelia (TSA), I love you all more than words can express. I hope I have shown you that you can achieve great things through perseverance.

Really though, I could not have done this without the support of my husband, Matt. We met in college, and when we went to our pre-Cana classes through the church, the priest told us we “wouldn’t have a pot to piss in” because we were both becoming teachers and marrying so young. Despite the priest’s less-than-desirable attitude, we continued our journey together. Our relationship gave me strength and support when I needed it most throughout this journey. I love you, Matt.
Thank you to my parents who sacrificed everything for us kids. I would not be where I am today without their love and support. Remember all those terrible grades I earned in high school? I do, too. Thank you for putting up with me and for teaching me that I don’t get to give up just because things get tough. I love you both so much.

There were bumps along the way on this journey to EdD. I actually “quit” the program a few years ago, and I did so at my grandparents’ house in Kansas City. I distinctly remember sitting on their couch and crying to them that I was done; that I couldn’t do it anymore. When I decided to go back and finish, they were the first ones I told, and they continued to be the first ones I called with each milestone I hit in the process. Thank you for your love and support, Grandma and Grandpa. I love you both.

Last, but not least, I want to acknowledge the support of my dean, Marianne Hunnicutt, as well as the support of my colleagues. Thank you for leading the way.
DEDICATION

To those students who do not play the game of school well: we will do better for you
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................. ix
LIST OF FIGURES ................................................................................................................ x
LIST OF APPENDICES .......................................................................................................... xi

Chapter

1. INTRODUCTION ................................................................................................................. 1
   Theoretical Framework ........................................................................................................ 5
   Problem and Purpose Statements ...................................................................................... 6
   Research Questions ........................................................................................................... 7
   Significance of the Study .................................................................................................... 8
   Definitions .......................................................................................................................... 9
   Delimitations ...................................................................................................................... 10
   Methodology ..................................................................................................................... 10
   Organization of Dissertation ............................................................................................ 11

2. REVIEW OF LITERATURE ................................................................................................. 12
   Grades as a Form of Communication .............................................................................. 12
   Achievement vs. Behavior ............................................................................................... 17
   Progressive Perspective .................................................................................................... 18
   Accuracy ............................................................................................................................ 18
   Determining Student Grades in American K-12 Public Schools ..................................... 19
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grading Systems in American Higher Education Institutions</td>
<td>21</td>
</tr>
<tr>
<td>Assessment Practices of Instructors in Community Colleges</td>
<td>25</td>
</tr>
<tr>
<td>Investigating Grading Practices</td>
<td>28</td>
</tr>
<tr>
<td>Validity and Standards-Based Reporting - A Conceptual Framework</td>
<td>29</td>
</tr>
<tr>
<td>K-12 Teachers’ Views of Grading and Reporting Issues</td>
<td>37</td>
</tr>
<tr>
<td>Conclusion</td>
<td>39</td>
</tr>
<tr>
<td>3. RESEARCH DESIGN</td>
<td>41</td>
</tr>
<tr>
<td>Research Questions</td>
<td>41</td>
</tr>
<tr>
<td>Research Design</td>
<td>42</td>
</tr>
<tr>
<td>Community College and Participants</td>
<td>44</td>
</tr>
<tr>
<td>Data Collection</td>
<td>47</td>
</tr>
<tr>
<td>Survey Instrument</td>
<td>47</td>
</tr>
<tr>
<td>Interviews</td>
<td>51</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>52</td>
</tr>
<tr>
<td>Quantitative Analysis</td>
<td>52</td>
</tr>
<tr>
<td>Qualitative Analysis</td>
<td>57</td>
</tr>
<tr>
<td>Summary</td>
<td>59</td>
</tr>
<tr>
<td>4. RESULTS</td>
<td>61</td>
</tr>
<tr>
<td>Quantitative Analysis</td>
<td>61</td>
</tr>
<tr>
<td>Description of the Sample</td>
<td>61</td>
</tr>
<tr>
<td>Demographic Information</td>
<td>61</td>
</tr>
<tr>
<td>Descriptive Statistics for the Instrument Used in the Study</td>
<td>63</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Theme 3: Behaviors</td>
<td>91</td>
</tr>
<tr>
<td>Further Analysis</td>
<td>92</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>93</td>
</tr>
<tr>
<td>Research Question 4</td>
<td>94</td>
</tr>
<tr>
<td>Conclusion</td>
<td>95</td>
</tr>
<tr>
<td>5. OVERVIEW OF THE STUDY</td>
<td>97</td>
</tr>
<tr>
<td>Discussion of Major Findings</td>
<td>97</td>
</tr>
<tr>
<td>Impact of Academic Subdivision</td>
<td>97</td>
</tr>
<tr>
<td>Impact of Years in Education</td>
<td>98</td>
</tr>
<tr>
<td>Impact of Learning to Grade</td>
<td>100</td>
</tr>
<tr>
<td>Implications for Practice</td>
<td>101</td>
</tr>
<tr>
<td>Limitations of this Research</td>
<td>103</td>
</tr>
<tr>
<td>Opportunities for Further Research</td>
<td>103</td>
</tr>
<tr>
<td>Conclusion</td>
<td>105</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>107</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>113</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Messick’s Categories as They Pertain to Grading</td>
<td>32</td>
</tr>
<tr>
<td>2. Facets of Validity as a Progressive Matrix</td>
<td>33</td>
</tr>
<tr>
<td>3. Variables of the Study</td>
<td>37</td>
</tr>
<tr>
<td>4. Data Analysis and Research Questions</td>
<td>53</td>
</tr>
<tr>
<td>5. Dependent Variables for Research Questions 1 and 2; Latent Variable: Perception</td>
<td>55</td>
</tr>
<tr>
<td>6. Assumptions of MANOVA</td>
<td>56</td>
</tr>
<tr>
<td>7. Frequency for Nominal and Ordinal Variables</td>
<td>62</td>
</tr>
<tr>
<td>8. Summary Statistics for Interval and Ratio Variables</td>
<td>63</td>
</tr>
<tr>
<td>9. Eigenvalues, Percentage of Variance, and Cumulative Percentage for Factors for the Four-Item Variable Set</td>
<td>68</td>
</tr>
<tr>
<td>10. Factor Loadings from Exploratory Factor Analysis</td>
<td>68</td>
</tr>
<tr>
<td>11. Correlations Between Dependent Variables</td>
<td>71</td>
</tr>
<tr>
<td>12. MANOVA Results for Progressive Perspective, Communication, AchvVSBeh, and Accuracy by Division</td>
<td>72</td>
</tr>
<tr>
<td>13. Frequency for Nominal Variables – Faculty without A Degree in Education</td>
<td>75</td>
</tr>
<tr>
<td>14. Frequency for Nominal Variables – Faculty with A Degree in Education</td>
<td>75</td>
</tr>
<tr>
<td>15. Analysis of Variance for Purpose by EdDegree, Years ED_recode, Division, and Gender</td>
<td>78</td>
</tr>
<tr>
<td>16. Semi-Structured Interviewees</td>
<td>81</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Messick’s Theory of Validity as a Conceptual Framework for a Standards-Based Reporting Model</td>
<td>30</td>
</tr>
<tr>
<td>2. Mahalanobis Distance Scatterplot Testing Multivariate Normality</td>
<td>65</td>
</tr>
<tr>
<td>3. Scree Plot with the Kaiser Criterion</td>
<td>66</td>
</tr>
<tr>
<td>4. Chi-Square Q-Q Plot for Squared Mahalanobis Distances of Model Residuals to Test Multivariate Normality</td>
<td>70</td>
</tr>
<tr>
<td>5. Purpose of Grades</td>
<td>74</td>
</tr>
<tr>
<td>6. Q-Q Scatterplot for Normality of Residuals for the Regression Model</td>
<td>77</td>
</tr>
<tr>
<td>7. Studentized Residuals Plot for Outlier Detection</td>
<td>77</td>
</tr>
<tr>
<td>8. Means of Purpose by YearsED</td>
<td>79</td>
</tr>
<tr>
<td>9. Codes for How Full-Time Faculty Learned to Grade</td>
<td>84</td>
</tr>
<tr>
<td>10. Codes for Currently Part of Students’ Grades</td>
<td>87</td>
</tr>
<tr>
<td>11. Codes for Grading Components in an Ideal World</td>
<td>93</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. INSTITUTIONAL REVIEW BOARD APPROVAL</td>
<td>113</td>
</tr>
<tr>
<td>B. SURVEY PARTICIPANT CONSENT</td>
<td>116</td>
</tr>
<tr>
<td>C. INTERVIEW PARTICIPANT CONSENT</td>
<td>119</td>
</tr>
<tr>
<td>D. SURVEY INSTRUMENT</td>
<td>122</td>
</tr>
<tr>
<td>E. SUPPORT TO REPLICATE STUDY</td>
<td>126</td>
</tr>
<tr>
<td>F. SEMI-STRUCTURED INTERVIEW PROTOCOL</td>
<td>128</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Grading systems in early America were based largely on the European model, which focused on competition, prizes, and rank, and were used for internal communication among teachers and families (Schneider & Hutt, 2014). When compulsory education was implemented in the United States and teachers began to work with an increasing number of students, there was a need to develop a system that allowed teachers to quickly communicate with students and parents (Guskey, 2015). Thus, by the early 1900s, a single grading system was in place in most K-12 schools as well as in most institutions of higher education (Guskey). This also led to grades becoming a form of external communication, which opened the floodgates for policy makers and administrators to shape grades into a more legible form of communication (Schneider & Hutt, 2014).

The first letter grades in higher education can be traced back to Harvard in 1883 (Mathews, 2005). The letter itself was a “B,” but what that “B” represented was up for debate. For example, the “B” could mean “very little improvement,” but to others, it could mean “above average.” Beyond the grade at Harvard, the first actual grading system in higher education was developed at Mount Holyoke College in 1897 (Mount Holyoke, 2006). The system at Mount Holyoke College included the letters to represent meaning: A-Excellent, B-Good, C-Fair, D-Barely Passed, and E-Failed.
While options for grading systems have been explored since the late 1800s, there has been very little research done to understand what grades are meant to communicate (Guskey & Jung, 2006, Marzano, 2009, Wormeli, 2013). One way to think about grading was to think of it as just reporting student learning (Guskey, 2009), which should be fair and accurate (Wormeli, 2013). According to Green and Emerson (2007), grading was “the process by which the work was assigned some code – usually A through F or a percentage” (p. 4). According to Guskey and Bailey (2001), reporting student learning certifies attainment of learning goals, identifies where more work was needed, and offers a basis for improvement efforts. However, when examining current practices in American education, grading and reporting student learning are two very different constructs, and they are not always dependent on one another.

In considering these ideas, it becomes apparent that reporting student learning can occur without letter grades. However, the reverse was not true. Grades cannot and should not be reported without a connection to student learning regardless of how ingrained in American culture this practice has become (Guskey, 2015). The issue lies in the fact that most people are comfortable with seeing a single letter grade on a report card, and most are confident in what that letter may represent regarding a child’s academic performance (Wormeli, 2013). While this may be the traditional method of grading, and one that was common in schools, Messick (1994) argues that it was essential to understand that not all grades measure what they purport to measure.

When Guskey (2009) talks about grading, he delineates between two levels of education: elementary and secondary. He contends the distinction was important because grading expectations differ across those levels. For example, in elementary school, grades are seen “primarily as a way to communicate with parents, and [they] more often distinguished
achievement from behavior indicators” (p. 2). Whereas, “Secondary teachers based their grading practices on what they perceived would best prepare students for college or the work world, believed that grades helped teachers influence students’ effort and behavior, and were committed to the mathematic precision of grade calculations” (p. 2).

Guskey (2009; 2015) argues that the purpose for which teachers use grades can also vary greatly from one teacher to the next. While most educators strive for consistency within their classroom, beyond the classroom it was not known if there was any consistency in grading practices among teachers, across grade levels, or across disciplines (Wormeli, 2013). This statement was even true when teachers have common assessments given across a grade level or course. For instance, when using a 100-point scale to grade students, Wormeli contends there was an abundance of subjectivity between each of those 100 points. When considering the 100-point scale, unless each point was clearly delineated, a teacher’s decision to award points, was subjective. When those subjective grades are averaged together, the subjectivity and inaccuracy grow.

Schneider and Hutt (2014) explain that as grades became more prevalent in the United States, grades became more of an indicator of student ability and achievement, which led to grades communicating more powerfully. Each individual teacher’s understanding, level of training, and beliefs about grading, leads to how they determine student grades. Thus, you could have students who perform similarly on assessments, but receive different grades (Guskey, 2015). Considering the subjectivity, especially in higher education, where faculty members are much more autonomous, one must question why American education relies so heavily on a system that may inaccurately report student learning through letter grades because grades on individual assessments are determined subjectively by each individual instructor. While these
grades may be tied to course objectives, the numerical assignment to the grade was subjective, especially when there are 100 levels of performance, which was a traditional 100-point grading scale. When you take the average of all of the subjective grades on individual assignments, the subjectivity of the grade only grows. (Guskey, 2015). Each individual faculty member was able to determine student grades based on any criteria they deem important to the course. Ironically, grades may become even more important for students at this time because they may have a desire to be admitted to a particular degree program with a minimum Grade Point Average (GPA) requirement. Additionally, administrators may use grades as a data point to examine the performance of a particular program or division of the institution by aggregating grades given, which can add to the pressure that faculty members face, and, as a result, faculty may feel compelled to issue students certain grades to prevent scrutiny or dissatisfaction from administrators. Administrators may also use the aggregated grade data to evaluate programs or to make determinations about the overall effectiveness of particular departments, or they may examine grades of individual faculty members to see how many of each letter grade was issued, ultimately ensuring faculty achieve a specific grade distribution that has been deemed socially acceptable at that particular institution (Guskey, 2015). This idea of being socially acceptable extends beyond the visible grade used to communicate to the student and parents, and seeps into such practices as reducing the number of students underperforming for federally funded grants and programs. The practice for this particular study was focused on the even distribution of A’s, B’s and C’s, with few D’s and F’s, which was consistent in secondary as well as higher education (Schneider & Hutt, 2014).

Grades, while only one aspect of the American education system, can have a significant impact on a student’s overall success in their academic career (Wormeli, 2013). Since American
public schools rely so heavily on grades, it was important to understand the perception of a single letter (Guskey, 2009; Marzano, 2009; Reeves et al., 2017; Wormeli, 2013). Guskey conducted a study 556 K-12 teachers to examine perceptions of grades. The findings revealed numerous inconsistencies among the different groups of stakeholders in what they intend to communicate with grades as well as their perception of grades (Guskey). Knowing this, it was important to analyze the perception of the individuals who assign a single letter to represent student learning. The findings from Guskey’s study initiated a conversation for K-12 education, but this type of research has not been conducted in higher education. According to Schinske and Tanner (2014), more research on grading in higher education was needed. For the purpose of this study, a large Midwestern community college will be the focal point. A community college student’s grades gain importance when students are trying to transfer from a community college to a four-year institution to complete a baccalaureate degree. The importance of grades in this situation was the result of certain grade point averages being required for admission to institutions of higher learning and for individual programs within those institutions. Because of this context, it was necessary to understand grading in the community college by studying the perceptions of grades of full-time instructors.

Theoretical Framework

The theoretical framework for this study was anchored in Messick’s (1994) theory of construct validity. This theory is used in this study to form a framework to examine the full-time faculty members’ perceptions of course grades. Messick’s theory combines scientific inquiry and rational argument for scores, both of which are used in traditional as well as alternative forms of grading in American public schools. To better understand this theory, it was central to define the
term *construct*. According to Brown (2000), “a construct was an attribute, proficiency, ability, or skill that happens in the human brain and was defined by established theories” (p. 9). Brown describes construct validity as an “experimental demonstration that a test was measuring the construct it claims to be measuring” (p. 9). The concept of validity encompasses the meaning of a score as well as the implied meaning from the interpretation of the said score, all of which play a crucial role in examining teachers’ grading practices (Brookhart, 1993; Messick, 1989). In this study, I examined faculty members’ perceptions of grades through the lens of validity. Using this lens indicates that the meaning of a grade, a “C” for example, as well as the implied meaning of the grade, or what the faculty member was implying through the grade of a “C” was examined. Examining how student learning was reported through the lens of construct validity helped to determine if our current, traditional method of grading was a valid way of reporting student learning, in this case, at the community college level. Messick’s theory of validity will be described in more detail in Chapter 2.

**Problem and Purpose Statements**

The most common form of grading presently in the United States was using averages to calculate a student’s grade (Abou-Sayf, 1996). In a traditional grading model, these averages typically include formative assessments, summative assessments, life skills and everything in between, but even though this practice was deeply rooted in tradition, it does not allow student learning to be reported accurately (Guskey, 2015; Reeves et al., 2017; Schneider & Hutt, 2014; Wormeli, 2013). Additionally, Guskey (2015) states “even school leaders who have some knowledge of effective grading policies and practices typically find it difficult to challenge these long-held and deeply entrenched grading traditions” (p. 10). This statement, although based on
experience, sets the stage for the need to further investigate why grades are given, what grades mean, and which ways of grading are most effective. Exploring effective grading policies and practices for implementing them have proven quite challenging in the United States (Guskey; Wormeli, 2013).

Studies to examine perceptions of teachers, parents, and students in K-12 in regard to grading have been done by researchers in the field of assessment (e.g., Brookhart & Nitko, 2008; Marzano, 2009; Messick, 1994), but little research exists on the topic in higher education (Guskey, et al., 2015). One of the main reasons K-12 teachers structure grades the way they do was to prepare students for college (Guskey, 2009). With this idea in mind, it was logical to wonder what higher education faculty intended to communicate through their grades and their perceptions of reporting student learning. Given the dearth of research specific to higher education on this topic, a study with this focus seems timely. Therefore, the purpose of this study was to examine community college full-time faculty members’ perceptions of grading and reporting student learning via course grades.

Research Questions

This study was guided by the following questions:

1. What differences exist in community college full-time faculty perceptions of grading?
   a. How are perceptions different by academic subdivision?
      • Null Hypothesis: H01a: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$
      • Alternative Hypothesis: H11a: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \neq \mu_7$
   b. How are perceptions different between faculty with and without an education degree?
      • Null Hypothesis: H01c: $\mu_1 = \mu_2$
• Alternative Hypothesis: H11c: \( \mu_1 \neq \mu_2 \)

c. How are perceptions different by gender?

• Null Hypothesis: H01c: \( \mu_1 = \mu_2 \)

• Alternative Hypothesis: H11c: \( \mu_1 \neq \mu_2 \)

2. How does the purpose of grades differ among community college full-time faculty?

3. Null Hypothesis: H02: \( \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 \)

4. Alternative Hypothesis: H12: \( \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \)

3. What criteria do community college full-time faculty members use to determine student grades?

4. What past experiences do community college full-time faculty members attribute to their learning how to determine grades?

Research questions one and two were analyzed using quantitative methodology, while questions three and four were analyzed using qualitative methodology.

Significance of the Study

Research (Guskey, 2015; Reeves et al., 2017; Wormeli, 2013) suggests that reliance on a single letter grade was inadequate because it does not accurately represent student learning, and most importantly, those letter grades may not have the same meaning when given out by different faculty (Guskey, 2009). Guskey (2009) has studied the perceptions of teachers, students, and parents in K-12 to better understand the meaning of grades. This study extended Guskey’s research into higher education, specifically with community college full-time faculty. At community colleges, where most students transfer to another institution of higher education, reliability and accuracy in reporting student learning should be of the utmost concern to students,
teachers, faculty members, parents, administrators, board members, and community members. The findings from this study will help to understand how faculty report student learning and will begin critical dialogue within the community college community about the reliance on a single letter grade.

Definitions

The following definitions are important to this study:

*Formative assessment* (assessment for learning) is a way of providing tools to teachers and students to improve and learn more effectively (Dixson & Worrell, 2016).

*Process grades* are those grades that represent how the student arrived at the product grade. Process grades may include grades for life skills; a student’s effort, behavior, or work habits; as well as grades on homework or quizzes (Guskey & Bailey, 2010).

*Product grades* are those grades that refer to achievement or level of performance on summative assessments, such as portfolio assessments, examinations, or reports (Guskey & Bailey, 2010).

*Progress grades* report how far a student has come, or rather how much the student gained from learning experiences (Guskey & Bailey, 2010).

*Standards-based grading* is a system that assesses and reports student learning based on a set of standards and allows students to progress ahead in the standards or go back and review based on the students’ competency (Heflebower, Hoegh, & Marzano, 2014).

*Standards-referenced grading* is a system that gives students feedback regarding their proficiency on a set of standards, which the school uses to report the students’ performance, but
this system does not move students ahead or ask them to move backward based on their competency (Heflebower, et al., 2015).

*Summative assessments* (assessment of learning) are generally high-stakes assessments that are almost always graded, typically occur less frequently, and are always at the end of a segment of instruction (Dixson & Worrell, 2016).

**Delimitations**

This study was limited to one community college in the Midwest. Although there are many community colleges in the Midwest, “TSA” was one of the largest and offers the most variety in degrees and certificates. To preserve anonymity of the institution and of the faculty members involved in the study, a pseudonym, TSA, will be used. Survey responses were limited to full-time faculty at TSA, which has an accessible population of 305 people. To provide a more consistent population, adjunct faculty, who account for over 1,300 instructors, were not surveyed.

**Methodology**

This was a pragmatic parallel mixed methods study which utilized survey methodology as well as interviews. This type of study aimed to provide answers to the above research questions by asking participants the quantitative questions, while at the same time asking the qualitative questions, to be able to corroborate the findings in each of the quantitative and qualitative portions of the study. The participants were full-time faculty members at TSA, a large community college in the Midwest and the second largest provider of undergraduate coursework in their state. Using this group of participants provided I the opportunity to investigate a variety
of faculty perspectives about grading and reporting student learning. Data for the survey was collected through an online survey using Qualtrics. The numerical data was analyzed using inferential statistics to answer each research question. The qualitative data was collected through semi-structured interviews and coding was used to identify themes.

Organization of Dissertation

This study will be presented in five chapters. Chapter 1 presents an introduction and overview of the study. This chapter includes a statement of the research problem, the purpose and significance of the study, and the research questions that were answered. Chapter 2 presents a review of the literature relevant to reporting student learning. Chapter 3 includes a description of the methods to be used in this study. In Chapter 4, the results of the study are presented and summarized. Chapter 5 includes a discussion of the findings, recommendations for practice, and suggestions for future research.
CHAPTER 2
REVIEW OF LITERATURE

This literature review will address aspects of grading in both K-12 as well as higher education. It will begin with examining grades as a form of communication and will then move to looking at grading systems in American public schools, and then on to American higher education institutions. Research regarding assessment practices of instructors in higher education will follow, along with research regarding the investigation of grading practices in higher education settings. Finally, this literature review will conclude with an examination of the research regarding teachers’ views of grading.

Grades as a Form of Communication

What the teacher intends to convey with the grade compared to how it was interpreted by parents was a challenge to traditional grading models in American public schools. In a study of fourth grade students and their parents in 16 Iowa schools, Waltman and Frisbie (1994) found great variability among how parents interpret their students’ grades and “an intolerable level of inconsistency between teacher and parents in the way grades from a given classroom are interpreted” (p. 223). The study analyzed three main areas: parents’ perceptions of teacher’s intended meaning; grade distribution; and beliefs about what should be included in a report card grade. Interestingly, 96% of all parents surveyed responded that achievement should be included, while only 88% of teachers responded the same (Waltman & Frisbie, 1994). Since both of the percentages reported are high, it was logical that including achievement on a report card was
important. While these statistics speak to the discrepancy, they also shed light on the understanding of the word achievement by teachers and parents. Part of the argument in the discrepancy was that teachers, students, parents, and administrators are reliant on a single symbol to communicate numerous types of information. To further understand this discrepancy, one must understand codification and code. Codification was defined as “explicit descriptions or statements composed of words, diagrams or symbols” (Sadler, 2014, p. 274). These explicit descriptions can be content standards, professional standards, or any other descriptor. Sadler describes these codifications as “the primary tools for communicating, transferring and sharing standards knowledge among learners, academics, accreditation agencies, professional bodies and employers” (p. 275). To differentiate between codification and code, Sadler defines codes as “an alphanumeric symbol or phrase such as Pass/Fail; A, B, C, D, F; Distinction, Merit, Credit, Pass, Fail; or 5, 4, 3, 2, 1, 0” (p. 275). Interestingly, Sadler notes that codification alone cannot be interpreted by different people in the same way because the codification was set in a specific context. When you remove, or change, the context, the codification inherently changes as well. Sadler remarks, “Codifications therefore cannot ‘hold’ standards by serving as stable reference points for judging and reporting different levels of student achievement” (p. 275).

Swan et al. (2014) conducted an exploratory study regarding parents’ and students’ perceptions of standards-based and traditional report cards. There were 24 teachers from a mixed elementary and middle school, grades 3-6, that participated in the study. At the end of the 9-week marking period, parents/guardians of students in those classrooms were surveyed regarding their perception of traditional grades compared to the new standards-based report card (Swan et al., 2014). There were a total of 115 parents/guardians who participated. Swan et al. (2014) argued that even in schools where standards have been clearly articulated, the use of a single
grade to report student achievement leaves too much room for interpretation and error. Sadler proposes the need for finding common ground for reporting student learning without having to standardize “the curriculum, teaching approaches or testing procedures” (p. 274).

In American public schools, grades have commonly been used as a communication tool among teachers, parents, students and administrators to represent everything from academic achievement and behavior as well as to communicate growth. These grades are then compiled from each content area and conveyed to stakeholders in the form of a report card. The report card was a tool that has been utilized in American public schools for over a century, but this tool lacks consistency, accuracy, and clarity and can cause confusion on behalf of any or all of the stakeholders (Guskey & Bailey, 2010; Munoz & Guskey, 2015). Part of the confusion comes from a lack of understanding of what grades really mean, but another issue with reporting was that distribution of grades was deeply rooted in American public-school culture (Guskey, 2002).

When parents, teachers, and students were surveyed about how grades should be distributed, all groups reported that there should be more A’s and B’s in elementary school, with few students earning low or failing grades, but in middle and high school, perceptions of grade distributions change (Guskey, 2002). Even in middle and high school, teachers report that most students should still be receiving higher grades; students report that, ideally, there should be more high grades; and parents report that grades should resemble a normal distribution pattern (Guskey, 2002). These beliefs are the crux of the traditional American public school system, and they impact teachers’ grading practices, as teachers can be questioned if the distribution of their grades does not reflect parental expectations (Guskey, 2015). This, too, puts pressure on a single symbol to communicate to parents; provide feedback for students; select, identify, or group students; provide incentives; evaluate school programs; and/or indicate lack of effort and
responsibility (Guskey, 2002). In the same study, these six purposes of grading were ranked by teachers, parents, and students. Communicating to parents received the highest ranking by all three groups of participants, which was a clear indicator that grades are used as a communication tool by stakeholders, although what that symbol communicates was a question that remains to be answered. Take, for example, Guskey’s 2009 study on teachers’ views of grading and reporting. In elementary schools, 90% of teachers believed that the primary purpose of report cards was to communicate with parents, while only 27% of secondary teachers perceived grades in the same respect (Guskey, 2009). While the instrument used in Guskey’s exploratory study (2009) to measure participants’ responses was grounded in research from a variety of sources, the participants were selected based on their involvement in professional development on performance assessments. This pool of participants included those from a northern state as well as a midwestern state, but demographics of the school districts were not provided for the study.

Friedman and Frisbie (1995) set out to explore the “meaningfulness of the interpretations parents might make about their child’s performance using report card information” (p. 6). Since the perception of grades was that they communicate, the authors examined the validity of making inferences on student achievement based on a symbol on a report card. Parents in American public schools draw their own conclusions about their child’s progress based on report card grades (Friedman & Frisbie, 1995). This practice was dangerous, as those inferences could potentially be the result of inadequate grading and assessment practices by the teacher rather than a lack of learning or understanding by the student. Take, for instance, a student who behaves poorly in class. If that student brings home a report card grade of C for a content area, does that symbol strictly represent the student’s content knowledge or does it also contain elements of the student’s behavior? Being unsure of the answer leads one to question the validity of the grade
and the validity of the inferences one draws from such a symbol. Looking critically at this issue, it can be observed that the grounds for the disconnected interpretations come from the poor practice of combining achievement grades with other soft skills such as effort, behavior, and improvement (Friedman & Frisbie, 1995). Understanding this disconnect, one should note that best practice in assessment and grading indicates that academic achievement and life skills should be separated on a report card (Guskey & Bailey, 2010; Munoz & Guskey, 2015; Wormeli, 2013). Finding an effective way to communicate this information without overwhelming teachers, parents, or students remains a challenge in moving away from traditional grading practices.

Another aspect of current grading practice in American public schools was that the perception of grades changes dramatically between elementary and secondary school teachers. In a quantitative study, Guskey (2009) found that elementary teachers felt grades were a method of communication and that achievement and behavior were separated in reporting grades. In the same exploratory study, Guskey also found that the secondary school teachers did not separate achievement from behavior and that the teachers believed grades influenced student behavior. More troubling, the secondary school teachers did not base grades on achievement, but instead on their perception of what would most benefit students in college or in their career. With this line of logic, each teacher was free to grade as they saw most beneficial for students, which could lead to significant variability in the validity of grades. Most troubling, secondary school teachers “were committed to the mathematical precision of grade calculations” (Guskey, 2009, p. 1). While mathematical calculation was precise in many instances, it cannot be utilized to determine grades because the grades themselves are the result of professional judgment of the
teacher and of indirect measurement of student learning. Basing grades on an individual teacher’s perception of grades opens the door for inaccurate reporting.

According to Guskey and Bailey (2010) and Munoz and Guskey (2015), it was possible to report student learning accurately using product, process, and progress criteria. Product criteria are used to communicate student achievement of content standards through the use of summative assessments; process criteria represent how the students got to where they are and may include responsibility, effort, work habits, quizzes, formative assessments, and homework; progress criteria are representative of student gain or educational growth (Munoz & Guskey, 2015). Keeping these three types of grades in mind, it was easy to see how inadequate the current method of grading and reporting in American public schools really was. In a traditional grading model, a single letter, A for example, was utilized to communicate a wide variety of information and that information was delineated in the product, process, and progress grades as described by Munoz and Guskey (2015).

Achievement vs. Behavior

Another form of communication that grades are used for is achievement and behavior. According to Guskey (2009), grades have a powerful influence on students, even though their psychometric properties are still considered to be questionable. The grades students receive have long-lasting effects on their persistence in school, as well as on their motivation to learn (Brookhart, 2004), yet psychometrically appropriate ways to communicate grades have yet to be widely adopted. Guskey (2009) also says that grades should ideally be used to communicate information to students regarding how they can improve, not used as a punishment for poor behavior. Reeves et al. (2017) state that some of the commonly used grading practices threaten
the emotional wellbeing as well as academic outcomes of students. Reeves et al. (2017) describe behaviors for grading such items as attendance, tardiness, inappropriate conduct, and turning work in late. On the other hand, grades can be based on academic achievement, and used to facilitate academic improvement in the student.

Progressive Perspective

Progressive grading practices include ideas such as teaching without grades, prioritizing the most recent evidence when determining grades, and reporting behavioral aspects of grades separately. One of the main reasons that grading was still such a prevalent issue in today’s society was because of teachers’ lack of formal training on grading and reporting (Guskey 2009). Most often, teachers base their grading practices on their own experiences, simply because of a lack of training, which can perpetuate inappropriate and sometimes damaging grading practices (Guskey, 2009; Reeves et al., 2017). Similarly, most teachers have little knowledge regarding the various grading methods, as well as their advantages and disadvantages, so teachers grade as they, themselves, were graded, ignoring a progressive perspective (Brookhart & Nitko, 2008; Guskey, 2009; Stiggins, 1993, 2008). There is a significant need for more research to find the experiences that lead to a progressive perspective with regards to grading and reporting practices.

Accuracy

Research shows that within the same school, and even within the same grade level, grading practices vary greatly (Cizek et al., 1996; Guskey, 2009; McMillan et al., 2002). Research further shows that even when teachers are given guidelines for grading, there was still great variance in the grading practices of individual teachers (Brookhart, 1994; Guskey, 2009;
McMillan, 2001). With all of this variance in grading practices, the idea of accurately reporting grades becomes a necessary idea to discuss. Further research must be done to determine how to more accurately report student learning. Sadly, there are currently no recommendations for more accurate grading in higher education, although at the K-12 level, there is a significant movement to incorporate standards-based grading practices. Standards-based grading practices are designed to assess students based on the standards and separates out behaviors from the content of the class. Research on this practice is still ongoing.

Communication, Achievement vs. Behavior, Progressive Perspective, and Accuracy are all critical components to explaining grading practices, hence their importance in this study. Each of these components will help to explain a perspective on grading, and the overall picture provided will become the construct of grading.

Determining Student Grades in American K-12 Public Schools

Lichty and Retallick (2017) conducted a descriptive survey research study of 236 high school agricultural educators in Iowa. One finding of the study was that there are numerous categories that comprise grades, including, examinations, quizzes, presentations, projects, homework, attendance, portfolios, participation, attitude, effort, and progress made (Lichty & Retallick, 2017). Although there was significant variance in what comprises a grade, there was some agreement among educators that the purpose of grades was for feedback about student achievement (Austin & McCann, 1992; Lichty & Retallick, 2017; Marzano, 2000). The Lichty and Retallick study (2017) found that of the 157 educators who responded to the survey, there was the most agreement that grades should reflect student learning. Additionally, respondents
agreed that a grade should be reflective of an individual’s achievement (Lichty & Retallick, 2017).

The study utilized a 5-point, Likert-type scale for respondents to rate their agreement with statements regarding traditional grading practices and the researchers then analyzed the responses to find that “academic achievement should be the primary basis for grades” (Lichty & Retallick, p. 14). Interestingly, the same study found that when respondents were asked to identify their actual grading practices, there was much more than just academic achievement involved in the practice. According to Lichty and Retallick (2017), knowledge, effort, responsibility, and attendance were all among the most common responses to the survey questions that asked the respondents to identify components of their actual grading practices.

In a quasi-experimental study conducted by Thiele (2018), research questions sought to learn about the impact grading policies have on middle school students’ mindsets. While the focus of the study was on mindset, the grading practices discussed in the study are of particular interest. Thiele (2018) aligns traditional grading practices with a fixed mindset, while standards-based grading practices are aligned with a growth mindset. The study explains that very few differences were found in the perceptions of grading in the two different types of grading practices being researched (Thiele, 2018). Following that line of logic, I also noted a “misalignment of students’ perceptions of the grading practices with the schools’ written grading policy” (Thiele, p. 5). While a clear understanding of grading practices was not defined in this study, I agree with several other researchers in this field to effectively implement grading practices, a clear purpose for grades must be defined (Guskey, 2015; Marzano, 2000; Thiele, 2018).
A quantitative study of higher education faculty conducted by Abou-Sayf (1996) investigated four different grading approaches: number grades, strict 5-point grades, judgmental 5-point grades, and judgmental 13-point grades. According to Abou-Sayf, these different grading scales are defined as follows: number grades – proportion of possible points that were actually earned by the student; strict 5-point grades – assigning point values to the letters, i.e. A = 4, B = 3, C = 2, D = 1, F = 0 and raw score intervals correspond to percentage score intervals, such as 90%, 80%; judgmental 5-point grades – the same letters and weights as the previous method, but instead of the cutoffs, the score distribution would determine the grade in an attempt to approach normality; judgmental 13-point scale – similar to the previous process, but this scale also includes + and - for each grade, A-D. It is important to note that according to Billet (1932), as of almost a century ago, more than 100 different grading procedures had been identified, though most higher education faculty use one of the following three procedures: number grades, which was both the most reliable and the most valid method, followed by the 13-point grades, and then the strict 5-point grades, which were the least reliable and valid. The results of this study have since been challenged, as Guskey and Bailey (2001) describe how adding plusses and minuses to grades actually diminishes the statistical reliability of the grade. It was important to note that there has been a resurgence in the use of percentage grades, thanks to technology and online gradebooks (Guskey, 2013). Also of importance was the understanding that grading scales in the early 20th century had an average of 50, with grades higher than 75 and lower than 25 being rare (Guskey, 2013). This was very different than modern grading, which typically sets the average at 75, with only grades above 60 as passing, which dramatically increases a student’s possibility for
failure, and one that was negatively skewed (Guskey, 2013). According to Guskey, “A student was statistically much more likely to be misclassified as performing at the 85-percent level when his true achievement was at the 90% level (a difference of five percentage categories) than he was of being misclassified as scoring at an Average level when his true achievement was at an Excellent level” (p. 3).

In a recent study conducted of grading practices in higher education, Norton et al. (2020) conducted a mixed methods study to investigate student and faculty perceptions of standards-based grading. While standards-based grading was considered a nontraditional grading practice, the components of standards-based grading contain some of the same categories as traditional grading practices (Guskey, 2015; Norton et al. 2020). Is note that even in this study, where the focus was on the perception of standards-based grading, traditional grading practices vary among faculty (Norton et al., 2020).

Norton et al. (2020) collected survey responses from 43 students in a clinical nursing capstone course. The reason this particular course was chosen was because of the lab component for completion of the nursing program. Based on the survey and subsequent semi-structured interviews, the researchers concluded that there were numerous positive themes regarding standards-based grading, many of which align with what other studies on perceptions of grading outline, including, measuring student achievement (Norton et al., 2020). In the portion of the study that addressed faculty perceptions, the researchers found themes regarding support in achieving student learning outcomes, accountability for the students, students’ control for their grade, motivation for students, creation of clear course and assessment grading expectations, and increased transparency between the student and faculty members. When analyzing the student
perceptions, themes found were similar to those from the faculty, but also included minimizing conflict with faculty member and promoting consistency in grading.

Kleinman et al. (2018) examined hundreds of colleges and universities in the northeastern United States to find out the percentage of college and universities that use full grade, A-F grading systems and to find out if there were identifiable demographic features and descriptors that led to specific grading systems. Data from 620 schools was gathered and analyzed in SPSS with the following measures: State, Enrollment, Degree type, Religion, and Grading scheme (Kleinman et al., 2018). Religion was utilized in this study because it was another way to group students. The study was trying to determine identifiable demographic features, so using religion as one of them appears to be logical. Using a Chi-Square test, the researchers found that there was a significant difference between the grading schemes in 2-year and 4-year schools, with 4-year schools using plus/minus grading more than 2-year schools (Kleinman et al., 2018). The researchers noted that 71% of schools studied that had a full grade A-F grading scheme were community colleges (Kleinman et al., 2018).

The three main reasons universities gave for utilizing a plus/minus grading scheme and those were: 1) better differentiate between students, 2) to help motivate students, and 3) to combat the ever-growing issue of grade inflation (Kleinman et al., 2018). Kleinman et al. (2018) says that “regardless of how one looks at evaluating and reflecting student work, in the final analysis, it remains a subjective estimation of the level of expertise each a student displays of a given curriculum” (p. 31). Kleinman et al. (2018) concluded that although higher education institutions are trying to move toward a more standardized model of grading, it may not be necessary because there was not much empirical evidence to show that a plus/minus grading
scheme was better than any others. The researchers say that this study lays the groundwork to explore grading systems in higher education across the United States (Kleinman et al., 2018).

Looking more at different grading systems, a quantitative study using archived performance data, sought to find out “how combining multiple grading practices and policies designed to promote excellence, differentiate among high-performing students, or prevent grade inflation at an elite institution affects GPA outcomes for low-performing students” (Wetzler, 2019, p. 1). At the institution of higher learning used for the study, there was a restricted percentage range for a “D” grade, 67.0-69.9%. Wetzler (2019) found that the restricted D range, in conjunction with a system that does not allow for plus/minus, heavily penalized low-performing students. At the opposite side of the spectrum, the restricted percentage for the “D” allows for differentiation among high achievers, since the differentiation among high achievers was then a smaller percentage (Wetzler, 2019). The study goes on to discuss the importance of having a good understanding of what letter grades represent regarding student achievement and that explanations of why cut-scores have been set where they are (Wetzler, 2019).

Following the same lines as the study above, Wilamowsky et al. (2008) conducted a quantitative study on the effects of plus/minus grading on the GPA. Wilamowsky et al. (2008) say that the most common reason to have a grading system with more flexibility (plus/minus) was to be able to differentiate between the performance of students. Studies have suggested that introducing plus/minus grades reduces or eliminates grade inflation, but they fail to examine the negative impact on students’ GPAs over time (Wilamowsky et al., 2008). Interestingly, Wilamowsky et al. (2008) examined the impact of adding plus/minus grading to a student’s GPA and they noted that there was a significant drop after the first year of implementation, but in the second-year student, GPAs are back where they were before the grading system change. Clearly,
there was debate among researchers as to the effectiveness of plus/minus grading schemes, but also about the impact of such decisions on students.

Finally, a study conducted by Sadler (2005), conducted a review of the most common grading policies in higher education that lean toward being criteria-based. In this review, Sadler found that there was no common understanding or common implications for what criteria-based means to faculty members. Sadler also found that faculty members make judgements about the quality of student work, and that the judgements are subjective and largely hidden from the student (Sadler, 2005). The following section will explore assessment practices of instructors specifically at the community college, as they relate to the proposed study.

Assessment Practices of Instructors in Community Colleges

Lei (2008) conducted an exploratory quantitative study \((n = 400)\) in which he investigated community college faculty’s use of assessment techniques to measure student achievement and mastery. It was important to note that although this study was published in 2008, the research was conducted in 2003. The factors, or independent variables, used for the study were Faculty Status (full-time vs. part-time) and education level (Ph.D. vs. M.A.). The formal assessment techniques used in the one-shot survey were derived from the work of Popham (2000, 2002) and include tests, quizzes, participation/discussion, classwork, homework assignments, portfolios, laboratory activities, cooperative learning, learning journals, research assignments, oral presentations, and peer- and self-assessments.

The survey instrument used in this study asked participants to respond on a 5-point scale. The survey was validated through a pilot study in which feedback was collected and used to make changes to the survey. Lei (2008) sampled two community colleges with uneven numbers
of faculty members, so a stratification sampling method was used. The response rate for the study was 45.8% and the highest response rate was noted in the first three weeks.

One of the findings of the study was that “full-time instructors learned their assessment techniques through a teaching and learning center, seminars or workshops, educational course/program, from former instructors, through personal experience, and from colleagues” (Lei, p. 396). Additionally, Lei (2008) found that full-time instructors emphasized the following: attendance/participation, quizzes, laboratory activities, cooperative learning, learning journals, research reports and projects, and online assessment. Lei (2008) reported significant differences in assessment techniques with regards to faculty status and educational level. For example, faculty members with doctoral degrees put more emphasis on lab activities and research projects than those faculty members without a doctoral degree. Lei (2008) also found that course grades “do not always reflect specific abilities of students” (p.400). In future research, attention would need to be given to the different types of assessments included in the survey to ensure that they reflect best practices in assessment.

BoarerPitchford (2014) conducted an exploratory quantitative study in which community college faculty were surveyed regarding their calculation of a course grade. BoarerPitchford collected the data through a one-shot survey of faculty at two community colleges in California; the survey was validated through a previous study conducted by Lei (2008). The study examined 17 assessment techniques that were determined through the constructs of traditional assessment and authentic assessment. The final question in the survey asked the faculty to estimate the percentage of the course grade for each of the 17 techniques, with the total percent being 100. The study cited a number of factors as being relevant including, faculty status (part-time vs. full-time); degree earned (Ph.D. vs. M.A.); and teaching experience (less than 3 years, 3 to 7 years, 8
to 15 years, more than 15 years). The study explored these factors as independent variables in a series of one-way ANOVAs.

BoarerPitchford (2014) found that full-time faculty put more emphasis on authentic assessments such as research projects and learning journals for the course grade, which was consistent with the findings in Lei (2008). But what the researchers failed to ask was what those projects were designed to assess. Having students complete a research project unrelated to the course concepts or objectives was no more worthwhile to mastery of the course content than lower-level assessments that only ask students to memorize and recall information. This research study (BoarerPitchford (2014) also failed to examine why the instructors based their grades on these types of assessments.

In the study conducted by BoarerPitchford (2014), learning journals were the only assessment technique that was statistically significant between the M.A. and Ph.D. level. BoarerPitchford (2014) hypothesized that the results of her study indicated doctoral-level instructors are using a variety of innovative assessment practices that encourage critical thinking, but further research would need to be conducted to confirm this hypothesis. Another hypothesis derived from the results of BoarerPitchford’s study was that doctoral-level instructors valued constructivism more than instructors with a lower level of education. She also stated that instructors with less than three years of teaching experience are likely to use assessment techniques that promote construction of meaning, which include hands-on activities to build understanding. Examining these ideas and being able to identify instructors based on their tendency to use a constructivist approach in assessment could potentially lead to increased student achievement, so research should be done to predict if an instructor would use a constructivist approach in the classroom.
BoarerPitchford (2014) suggested future research to examine why instructors employ specific assessment techniques. BoarerPitchford also suggested adding a qualitative component to a study to highlight how grades are actually calculated. Based on this recommendation, the proposed study includes a qualitative component in order to ask clarifying questions and follow up on data gathered from the online survey. It was also worth noting that since participants responded by self-reporting how they weight their grades, the results of this study may be skewed in favor of responses the more experienced instructors think they should be doing rather than what was actually happening in the classroom.

BoarerPitchford (2014) builds on the work of Lei (2008), as evidenced by the use of the same survey instrument and expands on the use of assessment techniques. The research was consistent that assessment techniques must be related to the course objectives and that the type of assessment instructors use plays a role in how instructors determine grades (BoarerPitchford, 2014; Lei, 2008).

Investigating Grading Practices

Abou-Sayf (1996) conducted a study at a community college that compared the different grading methods and analyzed them for reliability and validity. The study took the grades earned by 159 students in nine business classes over three semesters and transferred those grades into five different grading systems: number grades, strict 5-point grades, judgmental 5-point grades, and judgmental 13-point grades. Abou-Sayf used the Classical Test Theory, which says that the most efficient test has average difficulty and maximum discrimination for the group tested.

Cronbach’s Alpha was used in two ways to analyze the data: internal consistency and reliability of combined total class score. The validity coefficient was used to correlate class grade
and self-reported cumulative GPA. The findings of this study indicate that the most valid type of grading was the number grading method.

The small sample size as a limitation to the study. A consideration in regard to this study should also be the fact that only one instructor was included in the study. Further research that involves different instructors, especially in different content areas during the same semester, would be beneficial (Abou-Sayf, 1996). Number grades, which are the portion of points earned by students, were found to be the most reliable and valid (Abou-Sayf). While this point reflects the findings of Abou-Sayf’s particular study, the findings cannot be generalized to other instructors or institutions because the result was only really saying that using number grades was more reliable and valid when the same instructor assesses the same assignments in the same course. What remains to be examined was what type of grading model was most reliable and valid when different instructors over different content areas are examined.

Abou-Sayf (1996) argues that more accurate grading should decrease grade inflation, but he failed to examine what goes into the grade itself. Abou-Sayf does note that the exception to his entire argument was the use of criterion-referenced tests that require a certain score to show mastery. From reading this study, one can see that Abou-Sayf (1996) recognizes that the reliability and validity of grades are too dependent on each individual instructor’s decisions.

Validity and Standards-Based Reporting - A Conceptual Framework

According to Messick (1994), a seminal theorist in validity, “validity was nothing less than an evaluative summary of both the evidence for and the actual as well as potential consequences of score interpretation and use” (p. 6). His theory of score validation combines scientific inquiry and rational argument for scores and further explains why multiple pieces of
evidence are necessary to accurately report student learning. The six aspects of construct validity, as defined by Messick (1994) are: content, substantive, structural, generalizability, external, and consequential. These six aspects “function as general validity criteria or standards for all educational and psychological measurement” (p. 3). Interestingly, the aspects of validity directly relate to standards-based reporting, including presenting evidence, the ability to generalize scores across populations, and the value implications of score interpretation. To accurately report student learning, the assessments, grading, and reporting methods must all be valid. Figure 1 illustrates the necessary components to make meaningful inferences about students’ grades.

Figure 1. Messick’s theory of validity as a conceptual framework for a standards-based reporting model.

Similar to many other authoritative voices in assessment, Messick (1994) contends that “content standards specify what students should know and be able to do; performance standards specify the level and quality of that knowledge and skills that was deemed acceptable” (p. 5). In
a standards-based grading model, it was critical to have content standards clearly defined and to have performance standards written out in student-friendly language. Messick maintains a position he previously took in a 1989 article, which states that “the relation between the evidence and the inferences drawn should determine the validation focus” (p. 8). Applying this theory of validity to grading, one can see that it was critical for teachers and parents to be able to make accurate inferences about their child’s achievement in school. Traditional grading models in American public schools do not allow this to happen because of the over reliance on one symbol to appropriately communicate far too much information among teachers, parents, and students (Guskey & Bailey, 2010).

The concept of validity encompasses the meaning of a score as well as the implied meaning from the interpretation of said score, all of which play a crucial role in examining teachers’ grading practices (Brookhart, 1993; Messick, 1989). The validity question was framed “in terms of both meaning and values by identifying two facets of validity: the intended function for the score (interpretation or use) and the source of justification (empirical evidence) or social consequences” (Brookhart, p. 124).

The theoretical framework for this study was Messick’s theory of validity. In a study conducted by Brookhart (1993), Messick’s (1989) four categories, Construct Validity (CV), Relevance/Utility (R/U), Value Implications (VI), and Social Consequences (SC), determine the questions regarding grading that could be asked through that particular lens (Table 1). These four categories were also used in Brookhart’s study to analyze the qualitative data and assign a score based on the comments in the open-ended response questions. The values in the open-ended response column signify the score that the response received from the two raters. The score the response received reflected the highest level indicated in the comments (Brookhart, 1993).
Brookhart (1993) states that these categories are examined in Brookhart’s (1993) study as she delineates these categories and examines teachers’ grading practices to determine if each has an impact on grading and reporting student learning, as shown in Table 2. In this study, Brookhart (1993) uses construct validity “when score interpretation was supported by empirical evidence” (p. 124). Then, the relevance and utility of the scores are considered along with the construct validity. An example of construct validity and relevance and utility used in the study can be seen when breaking down the questions in the study: “Construct validity – what does the grade mean per se?” and “Relevance/utility – what does the grade mean when it was assigned to a student” (p. 125)?
Table 2
Facets of Validity as a Progressive Matrix

<table>
<thead>
<tr>
<th>Evidential Basis</th>
<th>Test Interpretation</th>
<th>Test Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct Validity (CV)</td>
<td>CV + Relevance/Utility (R/U)</td>
<td></td>
</tr>
<tr>
<td>CV + Value Implications (VI)</td>
<td>CV + R/U + VI + Social consequences</td>
<td></td>
</tr>
</tbody>
</table>

Looking at the Consequential Basis then, Brookhart (1993) adds the component of Value Implications – “What does the grade mean when it was assigned to a student, and of what value was it” (p. 126)? It also adds Social Consequences – “What does the grade mean when it was assigned to a student, and of what value was it, and what will happen because of it (Brookhart, p. 126)? Brookhart (1993) applies Messick’s (1989) theory of validity to grading in that teachers’ grading practices reflect their appraisal of the values and consequences of a grade. According to Brookhart (1993), grading was differentiated from other types of educational measures because interpretation and use was more blurred, and the use of grades drives the interpretation of such. Similarly, teachers playing a dual role of advocate for the student and passing judgment on the student can actually cause conflict, in so much that teachers have to decide what impact their judgment (grade) will have on the student. Brookhart further discusses the application of Messick’s theory of validity to grading when she says that the importance between empirical and consequential sources of justification was salient in the grading discussion.

To examine the construct of grading and teachers’ dual role of being an advocate and making a judgment about a student, Brookhart (1993) conducted a study in which graduate students in a M.S. Ed. program at Duquesne University who were currently employed, certified
teachers, completed a survey that contained seven scenarios each with one multiple choice question and one open-ended question, which had been validated through a pilot study. The seven scenarios were about grading and came from the work of Manke and Lloyd (1990) that investigated achievement and non-achievement factors used in teachers’ grading practices. As a result of Manke and Lloyd’s work, the survey instrument for Brookhart’s study included three grading scenarios regarding effort not matching ability, two scenarios regarding missing work, and two scenarios regarding improvement. To gather qualitative data and to further explore teachers’ values and the meaning of grades, the open-ended question, “Why did you make this choice?” was asked after the completion of each multiple-choice question. Since the question was consistent in all of the scenarios, placement of the scenario within the survey could have had some impact on the responses given by the participant, especially if the survey took longer than 20 minutes to complete. Surveys that take longer than 20 minutes to complete typically have a lower response rate as well as diminished quality of responses (Dillman et al., 2009).

Unfortunately, it was unclear how long the survey took to complete, and it was unknown if the scenarios were randomized when given to participants or if they were always in the same order. It would also be worthwhile to know if there were any differences noted in the types of responses provided by participants given the placement in the survey. Should scenarios be part of future research, much consideration should be given to the length of the survey to ensure that survey fatigue does not occur.

Brookhart (1993) broke the participants into two groups: teachers with educational measurement training (n=40) and teachers without educational measurement training (n=44). Eighty percent of the participants were female, the median years of experience was five, and the breakdown of participants by grade level was as follows: grades K-4=32%; grades 5-8=30%;
grades 9-12=23%; other=16% (Brookhart). The article does not identify or give examples of positions included in the “other” category.

Basic descriptive statistics were calculated for each of the seven items from the survey. The Mann-Whitney U Test was used to investigate if the degree of concern for values and consequences differed for the two groups in this study: those with educational measurement training and those without it. The four assumptions of the Mann-Whitney U Test were 1) dependent variable was measured at the ordinal or continuous level; 2) independent variable consists of two independent, categorical groups; 3) independence of observation, with no participant being in more than one group; and 4) variables are not normally distributed, but the shape of the distributions was the same or similar. The Mann-Whitney U Test and Chi Square tests revealed no statistical significance by measurement training with regard to the level of thinking about grade intentions and use.

The constant comparative method was used to examine the qualitative data gathered from the open-ended questions in the survey (Brookhart, 1993). Using this technique, Brookhart used all of the responses from each of the four Messick categories and compared within them to determine themes. The two major themes found in this study were school consequences and consequences beyond school. Within the school consequences theme, changes in student effort and attitude were identified. Within the consequences beyond school theme, most were related to the effect on student’s self-esteem and confidence. Brookhart analyzed the data by combining CV + R/U to examine teachers’ meaning of a grade from research question one. Similarly, VI + SC was examined to determine the value judgments of teachers regarding grading. No statistical significance was found by grade level in the analysis of the data, so it was reported in aggregate. If this study were to be replicated in higher education, it would be interesting to see if the finding
that grade level does not matter holds true. The question could become, “If or to what extent does grade level not matter when considering the meaning of grades and the value judgments of teachers regarding grading?” One could compare the data collected in higher education to determine if higher education functions similarly to K-12 with regard to teachers’ meaning of grades and value judgments of teachers regarding grading.

Defining the construct grade was an area that lacks research, especially in higher education. According to Brookhart (1993), the construct grade was equated to student work based on performance and was used as compensation for work completed at a certain level. When examining CV + R/U levels and describing the construct grade, differences were noted between the group who had received educational measurement training and the group who did not (Brookhart, 1993). Those with educational measurement training were more likely to only look at performance when finding evidence to use with a grade (Brookhart, 1993).

The results of Brookhart’s study (1993) are consistent with the work of Stiggins (1993) that found/proposed measurement instruction alone was not enough to explain differences between recommended and actual grading practices. Having said that, educational measurement instruction does not change the fact that teachers are concerned with fairness in grading. Examining Brookhart’s findings further, it was clear that teachers look at social consequences for assigning a particular grade, which was often why they examine effort when assigning grades. If a low-performing student gives his/her best effort and was still going to receive an “F,” the teacher was more likely to give him/her a “D.” In cases in which a high-performing student does not put forth the effort, the teacher was more likely to give that student an “F” (Brookhart, 1993). Examining this logic in higher education could prove to be useful.
K-12 Teachers’ Views of Grading and Reporting Issues

Teachers who teach the same course at the same level often consider vastly different grading criteria to describe student performance (Cizek et al., 1996; McMillan et al., 2002). Few teachers have formal training on grading, and most teachers have little knowledge of various grading methods and their respective advantages and disadvantages (Brookhart & Nitko, 2008; Stiggins, 1993, 2008). Simply put, teachers grade based on their own experiences rather than considering best practices in assessment to shape their grading decisions. Thus, grading varies greatly by teacher. To examine these ideas, Guskey (2009) conducted an exploratory study to investigate differences in views of teachers on issues regarding grading and reporting student learning. The factors considered in Guskey’s study include the teaching context: subject and grade level (Table 3). The study sought to determine K-12 teachers’ perspectives on grading and reporting issues.

Table 3
Variables of the Study

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Correlated Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject area</td>
<td>Responses to the items in the survey</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>School level taught</td>
<td></td>
</tr>
</tbody>
</table>

For this study (n=807), a one-shot survey of teachers in a medium-sized Midwest school district whose student demographics were reflective of the national demographics with regard to ethnicity/race and economically disadvantaged homes was used. The response rate of the survey
was 70%, with the different grade levels represented as follows: elementary (K-5) = 46%; middle school = 28%; high school = 25% (Guskey, 2009). Guskey also reported that 74% of the respondents were female.

The selection item on the survey was designed based on the work of various researchers regarding the primary purpose of grades for report cards (Feldmesser, 1971; Frisbie & Waltman, 1992; Linn, 1983). The Likert-type items asked participants to indicate their agreement or disagreement with various statements on grading policies and practices. Descriptive statistics for each item were run to analyze the data received from the 513 respondents. The data were then analyzed by subject area, grade level, and gender (Guskey, 2009). These factors were the differences that were examined using MANOVA. When examining the means differences, gender and subject area were not statistically significant but level taught was. Two-way interactions were analyzed between subject area and gender, subject area and school level, gender and school level, but none were statistically significant. Similarly, a three-way interaction was analyzed using subject area, gender, and school level and it also was not statistically significant. As a result of this, a more parsimonious model was developed with school level as the independent variable. In this model, 11 of the 30 items were statistically significant, including the following:

- Teachers can teach without grades;
- Students can and do learn without grades;
- Giving zeroes for work missed or worked turned in late was a questionable practice;
- Taking credit away from students for infractions was a questionable practice;
- I give priority to the evidence related to the most important learning goals or standards when determining grades;
● I assign “I” for incomplete grades;
● I use grading scales instead of percentages;
● Grades have some value as rewards but no value as punishments;
● I involve parents in homework;
● All students in my classroom are given the same homework assignments; and
● I grade all homework. (p. 18)

Examining this trend in views of grading and reporting to see how it transpires in higher education would be significant because many K-12 teachers emphasize preparing students for college; however, it was unknown what was important to college faculty regarding grades. Conducting research to find a baseline for where college faculty are with regards to defining their grading practices was critical.

Conclusion

There have been a number of studies on grading conducted in very recent years, which was certainly a change, as most of the research prior to the last five years was done in the mid-1990s. Due to the work of researchers such as Dr. Thomas Guskey and Dr. Susan Brookhart, the interest in finding better ways to communicate student learning has grown significantly. In addition to the dated research, there was also a lack of research in the area of higher education and reporting student learning in a way that was valid.

This research study, which will be discussed further in chapter 3, aimed to investigate how instructors in higher education perceive grades, which may lead to being able to determine practices that could potentially lead to more valid reporting. Much of the way secondary school
teachers grade was to prepare students for college and the workplace, but understanding how college instructors grade and what grades are intended to mean in higher education was still unknown. Additionally, understanding college instructors’ perceptions of grades may allow for college instructors to shift grading practices to more accurately report student learning.
CHAPTER 3
RESEARCH DESIGN

The purpose of this study was to examine full-time community college faculty members’ perceptions of grading and reporting student learning. This study compared the perceptions of faculty members to see if differences exist in the perceptions of grading based on a number of factors; it sought to identify the criteria that full-time faculty use to determine student grades. Full-time faculty members were asked to identify the academic subdivision in which they teach, as well as if they have completed a degree in an education field, which includes education degrees at the baccalaureate, masters, or doctoral level. Responses were compared to determine if differences in perceptions of reporting student learning existed between those who have a degree in education and those who do not. In addition, gender was a factor that was used in determining possible differences.

This chapter includes the following sections: research questions, research design, participants, data collection, data analysis, limitations, and summary.

Research Questions

Again, this study was guided by the following questions:

1. What differences exist in community college full-time faculty perceptions of grading?

   a. How are perceptions different by academic subdivision?

      • Null Hypothesis: $H_{01a}: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$

      • Alternative Hypothesis: $H_{11a}: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \neq \mu_7$
b. How are perceptions different between faculty with and without an education degree?

- Null Hypothesis: \( H_{01c}: \mu_1 = \mu_2 \)
- Alternative Hypothesis: \( H_{11c}: \mu_1 \neq \mu_2 \)

c. How are perceptions different by gender?

- Null Hypothesis: \( H_{01c}: \mu_1 = \mu_2 \)
- Alternative Hypothesis: \( H_{11c}: \mu_1 \neq \mu_2 \)

2. How does the purpose of grades differ among community college full-time faculty?

- Null Hypothesis: \( H_{02}: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 \)
- Alternative Hypothesis: \( H_{12}: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \)

3. What criteria do community college full-time faculty use to determine student grades?

4. What past experiences do community college full-time faculty attribute to their learning how to determine grades?

Research Design

This pragmatic parallel mixed methods study was designed to examine full-time faculty members’ perceptions of course grades. Given the mixed methods approach, the study was comprised of two parts, a quantitative component and a qualitative component. I utilized an online survey to investigate perceptions of full-time community college faculty members regarding the reporting of student learning as well as semi-structured interviews to investigate how faculty members determine student grades and how they have been prepared to make such decisions. This methodology was chosen because the research questions that guided this study
were both quantitative and qualitative. In this study data was gathered to answer four research questions; research questions one and two were quantitative, and questions three and four were qualitative. This study was based substantially on a 2009 study of elementary and secondary level teachers conducted by Dr. Thomas Guskey; however, this study was conducted at the community college level and included interviews to be able to further analyze and interpret the data. In addition to the fact that the instrument was utilized by Dr. Guskey, the survey was also intentionally chosen to gather data from a large number of people in a relatively short amount of time (Mertens, 2015). Creswell (2014) describes this type of survey as a cross-sectional survey design, which was best to explore and measure current attitudes and practices.

I used a simple descriptive approach, which comes from analyzing data at one point in time, also known as a one-shot survey (Mertens, 2015). Conducting a study with a one-shot survey poses a threat to internal validity because the design consists of a single observation. Selection threat was an example of threat to internal validity in this study. Selection threat means that there will be a lack of generalizability from the sample to the population. Using a one-shot survey also means that the results from this study cannot be generalized externally, which was a specific example of a threat to external validity.

The qualitative portion of the study was utilized to be both complementary and as an expansion of the quantitative data gathered from the online survey (Creswell, 2014). Since the quantitative survey was a one-shot survey, it was important to use this qualitative methodology to corroborate findings in the quantitative data. Additionally, the interviews allowed me to analyze the data for themes. Since participants were from different academic divisions on campus, these themes were identified by analyzing interview data from faculty members across all academic subdivisions at TSA. The intent was to be able to expand upon the findings of the
quantitative data and then complement it using the themes found in the interviews, thus making the study stronger and more representative of the faculty members’ perceptions related to grading. Using this design allowed me to elaborate on and clarify findings from the quantitative data and also allowed for expansion of the breadth and range of the study (Creswell).

Community College and Participants

Across the country, there are over 1,000 community colleges of varying sizes. Community colleges that are smaller typically have less variety in the courses, certificates, and degrees they have to offer to students, with the opposite true for larger community colleges. With that logic in mind, trying to develop a baseline understanding of faculty members’ perceptions about student learning at a larger community college was preferred because of my desire to study as large a group of participants as possible. Outside of the state of California, the largest community college in the United States was TSA. TSA hosts approximately 28,000 students each semester. TSA has over 300 full-time faculty members and over 1,300 adjunct faculty members. Faculty members at TSA serve an important role in that they primarily teach general education courses, many of which have been articulated through the Illinois Articulation Initiative (IAI). This articulation initiative aims to prevent students from having to repeat courses at a four-year institution that they have already completed at the community college by having specific requirements and objectives articulated and agreed upon by an IAI panel (iTransfer, 2014). Students who complete an IAI course are guaranteed to have that course transfer to another Illinois institution and fulfill a particular course requirement (iTransfer). Since the objectives of the courses have been clearly articulated, the grading should also be examined to ensure that all faculty members are accurately reporting student learning.
A convenience sample was used for the quantitative portion of the study. Convenience sampling, although it may be the least desirable sampling strategy, is widely used in the field of education (Mertens, 2015). Although it is widely used, it is important to note that since a convenience sample was used, results cannot be generalized. Also, since convenience sampling was used for this study, there was a concern of selection threat to internal validity. TSA fits the criteria of a large community college, offering more than 90 certificate programs and degrees and nine associate’s degrees in 59 occupational and 45 transfer pre-baccalaureate programs. Thus, the vast fields of study represented by full-time faculty at TSA provide the opportunity to gather data from various teaching backgrounds and professional experiences. The participants were all full-time faculty members, both male and female, and ranged in age from 25 to 65 years old. All of the participants had a master’s degree or higher, as that was a hiring guideline at TSA. The population for this study was approximately 300 participants, and those participants came from all academic divisions of the college. I was a faculty member at TSA, so the email addresses were obtained through the listserv of full-time faculty members.

An email message was sent to all full-time faculty members informing them of the study. The online survey was sent to full-time faculty members at the start of the spring semester because they are back on campus due to contractual obligations, have access to technology, and have the greatest chance to respond at that time of year. The participants received a maximum of five emails with the information about participating in the study. Participants did not receive any emails after completing the survey.

Determining G*Power, to do a priori planning, was based on the following conditions: a moderate effect size, alpha =.05, power = .80. As a result, a MANOVA with four dependent variables needed a minimum of $n = 100$ participants. The MANOVA model needed a minimum
of 100 participants, but since I also conducted an exploratory factor analysis, a minimum of 150 participants were required. The four composite, latent dependent variables were Progressive Perspective, Communication, Achievement vs. Behavior, and Accuracy. A composite variable, “grading” was then created when they, the four latent dependent variables, were placed into the MANOVA model. These latent dependent variables were used to analyze the grading, the dependent variable against the independent variables.

To further enhance the study, research questions three and four were analyzed using individual, semi-structured interviews. The participants were selected using stratified purposeful sampling. This technique was a combination strategy that allowed me to select participants based on specific criteria and then to choose a sample from the cases (Mertens, 2015). The specific criteria was “Degree in Education.” I invited eight individuals to participate in the interviews, four with a degree in education, and four without a degree in education. The interviewees were further identified by choosing at least one member from each of the academic subdivisions to ensure that the themes were drawn from interviews with a broad range of faculty members. With the invited individuals’ consent, they participated in the interviews. These interviews were important in this study because the questions posed to the participants in the quantitative portion of the study were not open-ended, so the interviews allowed participants to explain their thinking and to provide in-depth information regarding their perceptions of course grading. The interviews also allowed for the analysis of themes, which provided significant insight in the perceptions of course grades among faculty members.
Data Collection

Survey Instrument

The online survey was hosted by Qualtrics and contained the consent to participate, one selection item, 29 Likert-style questions, and demographic questions (See Appendices B and C). This survey was developed by Dr. Thomas Guskey and was used in a study he conducted in 2009. The six options for the selection item in the survey were created based on the work of numerous researchers (see Airasian, 2001; Feldmesser, 1971; Frisbie & Waltman, 1992; Linn, 1983). As Guskey describes, “the remaining rating scale items asked teachers to indicate their agreement or disagreement with statements about a wide range of grading policies and practices” (p. 4). The four-point scale for the Likert-type questions range from “Strongly Disagree” to “Strongly Agree” (Guskey; Merriam, 2009). Dr. Guskey’s approval to use this survey for this study is noted in Appendix D. The survey instrument that was used in this context was altered from the original survey used by Guskey (2009) to include language that pertains specifically to higher education. For example, In question 1, the word “District” was changed to “College” to represent TSA. Also, the word “teacher” was replaced with “faculty member” to represent the higher education faculty. Question 11 from the original survey (Guskey, 2009) was eliminated because it was standard practice at TSA to give an I for incomplete course grades, therefore, the question was unnecessary. The changes to the survey were necessary because the educational environment in which the survey was being used for this study was different than where it was originally used. It was important to adjust the questions so that the data collected reflected practices in higher education.
At the end of the survey, faculty members were asked to enter demographic data regarding their educational background, gender, and teaching assignment (academic division), as well as to indicate their willingness to participate in an interview. I added item #29 on the survey to address the issue of perception in consistency in assigning student grades. The data was collected through Qualtrics and then uploaded into Intellectus Statistics for analysis. Once the data were received, I examined the reliability of that item using Cronbach’s alpha to determine if the item should be included. I also examined the entire survey for internal consistency using Cronbach’s Alpha. An exploratory factor analysis is intended to explore underlying factors of the variables, and was also completed to report alignment with the latent variables: Progressive Perspective, Communication, Achievement vs. Behavior, and Accuracy.

Dillman et al., (2009) outline recommendations for high response rates to surveys. Their recommendations were followed for this study to ensure the highest response rate possible. A pre-notice email was sent that described the purpose of the study. Next, an email was sent with a link to the study in Qualtrics. There were follow up emails sent to people who did not respond after two days and again after seven days of not responding. A final notice was sent to those who did not complete the survey after 11 days.

Based on the literature (Brookhart & Nitko, 2008; Cizek et al., 1996; McMillan et al., 2002; Stiggins, 1993, 2008), the independent variables that should be included in a study regarding grading at the community college are

- Faculty status (full-time vs. part-time)
- Degree earned (master’s vs. doctoral)
- Years of teaching experience
- Training in educational measurement (grading)
The literature describes differences between full-time and part-time faculty, thus a focus on full-time faculty allowed me to more deeply analyze the data. The study did not ask the participants the highest degree they have earned. This survey deviated from the suggested independent variable, and asked participants if they have a degree in education. The degree in education could have been at any level, including the baccalaureate level, and was not limited to a master’s or terminal degree in the field. This was significant in this study because a degree in education is not required to teach at the higher education level, but it could be a factor in explaining faculty members’ perceptions of grading. A question was included in the survey that asked participants to identify the number of years they have been teaching at the college level. It was important to specify at the college level because some faculty members have experience teaching in K-12 settings as well. Since the requirements for grading are much more rigid at the K-12 level (Guskey, 2015), only the years spent teaching at the college level will be considered for this study.

In the survey, participants were asked if they have ever received training in educational measurement (grading). This was one of the questions that was expanded upon in the interview portion of the study. I examined what type of training the participant had and what they learned from the experience. The inclusion of content taught was potentially an important variable because the data was examined to see if differences exist based on content area or between content areas. At the community college, there were not enough faculty members in each content area to make that worthwhile to examine; rather, it was more beneficial to examine the data through the lens of subdivisions. Those subdivisions at TSA include:

- Content taught
- Gender
- Arts, Culinary, and Hospitality
- Business and Applied Technology
- Liberal Arts
- Nursing and Health Sciences
- Social and Behavioral Science and the Library
- Science, Technology, Engineering and Math
- Counseling and Advising

Using these subdivisions provided large enough sample sizes and allowed me to examine trends based on those subdivisions.

An online survey was chosen for this study because the participants being studied (i.e., TSA full-time faculty) had access to technology via employer-issued hardware, which was an important consideration when only 79% of the population in North America have the same access (Mertens). The advantages of using this type of survey include convenient access, lower costs, and faster responses (Mertens, 2015). This type of survey also allowed me to investigate the issue of reporting student learning while allowing participants to remain anonymous, which was important when investigating a sensitive topic (Mertens).

At the conclusion of the quantitative survey, respondents were able to indicate their willingness to participate in an interview by entering their email when asked if they would like to participate in a follow up interview. I then took that pool of willing participants and selected four individuals with a degree in education and four individuals without a degree in education. I also ensured that there was at least one participant from each academic subdivision. I then reached out to the selected participants to schedule the interview.
Interviews

The semi-structured interviews were conducted by me via Zoom, and I utilized technology to record the interview. Using the record function in Zoom allowed me to transcribe the interview and create a transcript. The interview protocol can be found in Appendix F.

Participants provided consent to have the interview recorded. This mode was chosen because Zoom has been utilized for remote learning by full-time faculty, which allowed participants to feel more at ease during the interview. Each participant was interviewed individually, at a time agreed upon between me and the participant. Each interview lasted no longer than 30 minutes. I provided each participant with a Zoom link to utilize for the interview. I recorded each of the interviews so that I was able to focus on the “discussion” while it was unfolding.

The questions for the semi-structured interviews can be found in Appendix F. The questions were designed to answer RQ3 and RQ4. A semi-structured approach was being used because I wanted the participants to be able to expand upon their thoughts, and so that I could follow lines of thought as the participant was speaking. RQ3 seeks to explain the criteria faculty use to determine student grades, so that question was asked directly to participants. A follow-up question asking participants if they have always used the same category aimed to promote more discussion and more in-depth answers to arrive at a more substantial answer to the research question.

Questions 1 and 5 in the semi-structured interview were designed to answer RQ4. These questions asked the participant about the training they have received with regards to educational measurement (grading). If participants did engage in training, there was a series of questions to
find out more about the type of training and its impact on the participant and their grading practices. Question 6 was intentionally placed at the end of the survey to ask participants about their ideal situation for grading.

Data Analysis

This study was a mixed methods study utilizing inferential statistics and semi-structured interviews (Mertens, 2015; Field, 2013; Gravetter, et al., 2005). Table 4 shows how each research question was analyzed.

Quantitative Analysis

The intent of the quantitative portion of the study was to faculty members’ view of differences and purpose in grades, while the intent of the qualitative portion of the study was to explore course grades and experiences of faculty members leading to those course grades. RQ1 explains differences between groups. In RQ1a, there were seven groups, the seven academic subdivisions at TSA. These seven groups of full-time faculty members were used to explain differences in perceptions of grading by academic subdivision. RQ1b explained differences between two groups: those with a degree in education and those without a degree in education. RQ1c explained differences between two groups of people: males and females.

RQ2 explained purpose and consistency in grades. Thus, the hypotheses for RQ2 examined two groups of faculty members, those with a degree in an education field and those without. RQ2 explained purpose and consistency in grades by way of the two groups of people. Based on the literature, I hypothesized full-time faculty members with a degree in education will
Table 4

Data Analysis and Research Questions

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Hypotheses</th>
<th>Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What differences exist in community college full-time faculty perceptions about grading?</td>
<td></td>
<td>MANOVA</td>
</tr>
<tr>
<td>a. How are perceptions different by academic subdivision?</td>
<td>Null Hypothesis: H01a: $\mu_1 = \mu_2 = \mu_3 = \mu_4$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alternative Hypothesis: H11a: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$</td>
<td></td>
</tr>
<tr>
<td>b. How are perceptions different by faculty with and without an education degree?</td>
<td>Null Hypothesis: H01b: $\mu_1 = \mu_2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alternative Hypothesis: H11b: $\mu_1 \neq \mu_2$</td>
<td></td>
</tr>
<tr>
<td>c. How are perceptions different by gender?</td>
<td>Null Hypothesis: H01c: $\mu_1 = \mu_2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alternative Hypothesis: H11c: $\mu_1 \neq \mu_2$</td>
<td></td>
</tr>
<tr>
<td>2. How does the purpose of grades differ among community college full-time faculty?</td>
<td>Null Hypothesis: H02: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$</td>
<td>ANOVA</td>
</tr>
<tr>
<td></td>
<td>Alternative Hypothesis: H12: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6$</td>
<td></td>
</tr>
<tr>
<td>3. What criteria do community college full-time faculty use to determine student grades?</td>
<td>Interviews were coded and themes were developed and analyzed</td>
<td></td>
</tr>
<tr>
<td>4. What past experiences do community college full-time faculty attribute to their learning how to determine grades?</td>
<td>Interviews were coded and themes will be developed and analyzed</td>
<td></td>
</tr>
</tbody>
</table>
report at a greater rate that they believe the purpose of grading was to communicate with stakeholders as opposed to those without a degree in an education field. Full-time faculty members without a degree in an education field will report a belief of the purpose of grades as a way to provide incentives to learn as opposed to those with a degree in an education field. Full-time faculty members without a degree in an education field will report lower levels of consistency in grades as opposed to those without a degree in an education field.

The data from the Likert survey questions was analyzed using means, standard deviations, kurtosis, and skewness. A frequency distribution was utilized to show how many times each value occurred, which was helpful in analyzing trends in the distribution of scores (Field, 2013). Calculating means, which uses every data point, was utilized to create a summary of the data (Field; Gravetter et al., 2005). The standard deviations measured the spread of data around the mean and was used to analyze consistency of scores from the Likert survey questions (Field; Gravetter et al.). To further analyze the first research question, I used data from the Likert questions on the survey to calculate a MANOVA. Scales were created, and similar items were grouped (Table 5). A MANOVA was conducted to look for differences in the Progressive Perspective, Communication, Achievement vs. Behavior, and Accuracy, which were latent variables, and were also observable items from the survey. Then, the MANOVA model created a linear, composite variable, grading, that will be tested against the independent variables based on gender, education background, and subdivision at TSA.

Additionally, I examined the correlation between a grade and the implied meaning by the faculty member to determine if there was a relationship between the two ideas, the relationship’s
Table 5
Dependent Variables for Research Questions 1 and 2; Latent Variable: Perception

<table>
<thead>
<tr>
<th>Dependent Variable (Sub-constructs)</th>
<th>Survey Item (Observables)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Progressive Perspective</strong></td>
<td></td>
</tr>
<tr>
<td>1 – Teachers can teach without grades</td>
<td></td>
</tr>
<tr>
<td>2 – Students can and do learn without grades</td>
<td></td>
</tr>
<tr>
<td>7 – I give priority to the most recent evidence when determining grades</td>
<td></td>
</tr>
<tr>
<td>9 – I give priority to the evidence related to the most important learning goals or standards when determining grades</td>
<td></td>
</tr>
<tr>
<td>10 – I report behavioral aspects separately</td>
<td></td>
</tr>
<tr>
<td>11 – I use grading scales instead of percentages</td>
<td></td>
</tr>
<tr>
<td>12 – Grades have some value as rewards but no value as punishments</td>
<td></td>
</tr>
<tr>
<td>21 – My students know the criteria/objective to be assessed</td>
<td></td>
</tr>
<tr>
<td>22 – I give students the opportunity to take risks in my classroom</td>
<td></td>
</tr>
<tr>
<td>24 – I believe that study groups improve student achievement</td>
<td></td>
</tr>
<tr>
<td>27 – I believe the extent to which my homework was completed relies on how meaningful it was to students.</td>
<td></td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td></td>
</tr>
<tr>
<td>8 – I give priority to the most recent evidence when determining grades</td>
<td></td>
</tr>
<tr>
<td>16 – Assessments must be sources of information for student and faculty members</td>
<td></td>
</tr>
<tr>
<td>17 – When I give an assessment, I give corrective feedback</td>
<td></td>
</tr>
<tr>
<td>18 – When I give assessments, I give students a second chance to show improvement</td>
<td></td>
</tr>
<tr>
<td><strong>Achievement vs. Behavior</strong></td>
<td></td>
</tr>
<tr>
<td>23 – I believe homework increases student achievement</td>
<td></td>
</tr>
<tr>
<td>25 – All students in my classroom are given the same homework assignments</td>
<td></td>
</tr>
<tr>
<td>26 – I grade all homework</td>
<td></td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
</tr>
<tr>
<td>3 – Mathematical precision does not yield fairer or more objective grading</td>
<td></td>
</tr>
<tr>
<td>4 – Averaging to obtain a course grade was a questionable practice</td>
<td></td>
</tr>
<tr>
<td>5 – Giving zeroes for work missed or work turned in late was a questionable practice</td>
<td></td>
</tr>
<tr>
<td>6 – Taking credit away from students for infractions was a questionable practice</td>
<td></td>
</tr>
<tr>
<td>13 – Grading and reporting should always be done in reference to learning criteria – never on the curve</td>
<td></td>
</tr>
<tr>
<td>14 – High percentages are not the same as high standards</td>
<td></td>
</tr>
<tr>
<td>20 – I match assessments to what I am explicitly teaching (i.e., I don’t use trick questions, new formats, or unfamiliar material)</td>
<td></td>
</tr>
</tbody>
</table>
magnitude, and its directionality to make some type of judgment about criterion validity, which measures how well one measure predicts an outcome for another measure. This process was important to the study because it allowed me to determine if the grade appropriately signifies the meaning intended by the faculty member.

MANOVA was used to “compare groups on several outcome variables” (Field, 2013, p. 624). In the case of this study, the MANOVA compared groups of faculty members based on their response to the demographic questions on the survey and then compared the means differences of their responses. The assumptions of MANOVA include independence, random sampling, multivariate normality, and homogeneity of covariance matrices, as seen in Table 6. Meeting the assumptions of MANOVA are important to this study to ensure multivariate normality.

Table 6
Assumptions of MANOVA

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>Residuals should be statistically independent</td>
</tr>
<tr>
<td>Random Sampling</td>
<td>Data should be randomly sampled from the population of interest and measured at an interval level</td>
</tr>
<tr>
<td>Multivariate Normality</td>
<td>Residuals have multivariate normality</td>
</tr>
<tr>
<td>Homogeneity of Covariate Matrices</td>
<td>The population variance-covariance matrices of the different groups in the analysis are equal</td>
</tr>
</tbody>
</table>

(Field, 2013, p. 642)
Qualitative Analysis

To examine Research Questions 3 and 4, a qualitative approach was used in the form of semi-structured interviews. Utilizing semi-structured interviews allowed me to elaborate on and clarify findings from the quantitative portion of the study, which was called complementarity (Mertens, 2015). Utilizing this interviewing strategy allowed me to gain a more thorough understanding of the faculty members’ perceptions of course grades. The interviews were transcribed using the transcript function on Zoom. Using this technology, I had the interviews transcribed as they were being conducted. After the interviews were completed, I reviewed the transcripts for accuracy. For any discrepancies, I accessed the recording of the interview, and corrections were made. To ensure trustworthiness of the interview data, I conducted member checking by sending a copy of the transcript to interviewees so they could examine it to ensure that, what I captured was, in fact, their voice. If participants felt that there are any errors in the transcript, they were able to submit changes to me via tracked changes in a Word document. At that point, I reviewed the suggested changes, accepted or rejected them based on the recording, and then sent them back to the participant for a final review.

After ensuring the transcripts were accurate, I followed a coding process that was recommended by Saldaña (2015). I laid out printed interview transcripts that were double-spaced on the right two-thirds of the page, with a large margin on the left side, as I was lefthanded, to write notes and to code. Structural coding was utilized to analyze the interviews, which, according to Saldaña (2015), was an appropriate strategy for semi-structured protocols. I started with precoding by highlighting quotes or passages that were worthy of future attention. In order for a quote to be worthy of future attention, it needed to explain an idea related to the research
question. For example, when coding the transcripts for Research Question 3, any quotes related to criteria of grading were reviewed. If I felt the quote was significant, or that it carried meaning with regards to the research question, then the quote was highlighted. This was done because examining these quotes is a critical step in determining themes. I then focused on those quotes or passages that were identified in precoding and began to develop preliminary codes. These preliminary codes were based largely on the frequency that they were repeated by the interviewees. Additionally, I looked for key words that stood out as being closely related to ideas that had been identified in the review of literature on grading.

After the initial review of the quotes and transcripts, and the development of the codes, I then went back to analyze the codes for overall themes. There were seven initial codes: experience, being a parent, evolving practices, coursework, professional development, colleagues, and seeing through students’ eyes.

Once the codes were determined, I analyzed the codes for frequency to determine the number of times each individual participant mentioned a particular code. Once those numbers were identified, I then entered all of the codes into Wordle, at the frequency that they were said, to create a visual account of the data, which can be seen in figures 9, 10, and 11. Then, I then engaged in a second cycle of coding called pattern coding (Saldaña, 2015). Through the use of pattern coding, I pulled information from the first cycle of coding and created more meaningful units of analysis. To accomplish this, I reviewed the preliminary codes from the quotes and transcripts by laying out the printed transcripts with the highlighted quotes and codes listed in the margin, and then analyzed the codes to find themes. I found these themes by connecting ideas from the quotes and other important ideas found in the interviews. For each of the research questions, I took the codes and then reviewed the transcripts to develop a better understanding of
how the ideas were connected. Using that understanding, I then began to connect ideas, or codes. This deeper understanding was then used to take the codes that were found for each of the research questions, connect them to each other, and develop themes. Using the themes found through the different levels of coding, in conjunction with the quantitative data collected from the survey, I was able to answer the research questions. Further discussion of the process can be found in chapter 4.

Since this study was conducted in my workplace, it was important to reduce bias as much as possible. To do this, I provided participants with a participant consent form, which needed to be signed prior to participating in the survey. The consent form outlined confidentiality and also informed participants that they could withdraw from the survey at any time, without repercussions. Since I was a faculty member, I did not discuss the study, survey, or interview with participants outside of the official study protocol. Although I was a faculty member, and the participants in the study were also faculty members, there was no concern about being coerced into participating because I was not in an authority or supervisory role with regards to the participants.

Summary

This mixed methods study utilized an online survey as well as semi-structured interviews. The quantitative portion, the online survey, was used to gather data on perceptions of reporting student learning from full-time community college faculty members at TSA. Descriptive and inferential statistics was used to analyze the data collected from the survey. The qualitative portion of the study, the semi-structured interviews, gathered data on the criteria used to
determine grades as well as the experiences faculty members attribute to how they determine student grades. The next chapter will examine the findings from the survey and the interviews.
CHAPTER 4

RESULTS

Quantitative Analysis

Description of the Sample

The sample used for this study was the full-time faculty at TSA. The hiring guidelines for TSA state that an applicant must have a master’s degree to apply for the position, so all participants in the study have a master’s degree or higher. Participants were asked to input select demographic information, including gender, years in education, degree in education, and which subdivision they teach in at TSA. The sample size for this study was 133. The descriptive statistics for the demographic information are found in the subsequent paragraphs.

Demographic Information

The accessible population for the study was full-time faculty members at a community college. The study population used in this particular research was full-time faculty members at TSA Community College. Since the participants in this study were drawn from the accessible population, some generalizations could be made from the data, despite using a convenience sample.

Summary statistics were calculated for each interval and ratio variable. Frequencies and percentages were calculated for each nominal and ordinal variable. The most frequently observed category of Gender was Female \( n = 71, 53\% \). The most frequently observed category of
Education Degree \((EdDegree)\) was No Degree in Ed Field \((n = 80, 59\%)\). The most frequently observed category of Years in Education \((YearsED)\) was greater than 15 years \((n = 68, 50\%)\). The most frequently observed category of Division was Liberal Arts (LA) \((n = 30, 22\%)\).

Frequencies and percentages are presented in Table 7.

Table 7

Frequency for Nominal and Ordinal Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>(n)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>71</td>
<td>52.59</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>44.44</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>4</td>
<td>2.96</td>
</tr>
<tr>
<td>EdDegree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree in Ed Field</td>
<td>54</td>
<td>40.00</td>
</tr>
<tr>
<td>No Degree in Ed Field</td>
<td>80</td>
<td>59.26</td>
</tr>
<tr>
<td>YearsED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;15 years</td>
<td>68</td>
<td>50.37</td>
</tr>
<tr>
<td>8-15 years</td>
<td>50</td>
<td>37.04</td>
</tr>
<tr>
<td>&lt; 7 years</td>
<td>16</td>
<td>11.85</td>
</tr>
<tr>
<td>Division</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arts, Communication, and Hospitality</td>
<td>19</td>
<td>14.07</td>
</tr>
<tr>
<td>Social and Behavioral Sciences and the Library</td>
<td>28</td>
<td>20.74</td>
</tr>
<tr>
<td>Nursing and Health Sciences</td>
<td>16</td>
<td>11.85</td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>30</td>
<td>22.22</td>
</tr>
<tr>
<td>Science, Technology, Engineering and Math</td>
<td>23</td>
<td>17.04</td>
</tr>
<tr>
<td>Business and Applied Technology</td>
<td>15</td>
<td>11.11</td>
</tr>
<tr>
<td>Counseling and Advising</td>
<td>4</td>
<td>2.96</td>
</tr>
</tbody>
</table>

\textit{Note.} Due to rounding errors, percentages may not equal 100\%.
Descriptive Statistics for the Instrument Used in the Study

The observations for Progressive Perspective had an average of 2.96 ($SD = 0.34$, $SE_m = 0.03$, $Min = 1.64$, $Max = 3.91$, Skewness = -0.08, Kurtosis = 1.08). The observations for Communication had an average of 3.11 ($SD = 0.44$, $SE_m = 0.04$, $Min = 1.25$, $Max = 4.00$, Skewness = -0.28, Kurtosis = 1.33). The observations for Accuracy had an average of 2.87 ($SD = 0.41$, $SE_m = 0.04$, $Min = 2.00$, $Max = 4.00$, Skewness = 0.53, Kurtosis = 0.31). The observations for AchvVSBeh had an average of 3.21 ($SD = 0.53$, $SE_m = 0.05$, $Min = 1.33$, $Max = 4.00$, Skewness = -0.60, Kurtosis = 0.37). The summary statistics can be found in Table 8. The measures of central tendency, the measures of dispersion, the measures of variable location, and the measures of variable density are within the realm of acceptable, hence why I chose and executed these techniques and why the variables of the study are suitable for examination.

Table 8

Summary Statistics for Interval and Ratio Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SD$</th>
<th>$n$</th>
<th>$SE_m$</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive_Perspective</td>
<td>2.96</td>
<td>0.34</td>
<td>135</td>
<td>0.03</td>
<td>1.64</td>
<td>3.91</td>
<td>0.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Communication</td>
<td>3.11</td>
<td>0.44</td>
<td>134</td>
<td>0.04</td>
<td>1.25</td>
<td>4.00</td>
<td>-0.28</td>
<td>1.33</td>
</tr>
<tr>
<td>Accuracy</td>
<td>2.87</td>
<td>0.41</td>
<td>134</td>
<td>0.04</td>
<td>2.00</td>
<td>4.00</td>
<td>0.53</td>
<td>0.31</td>
</tr>
<tr>
<td>AchvVSBeh</td>
<td>3.21</td>
<td>0.53</td>
<td>135</td>
<td>0.05</td>
<td>1.33</td>
<td>4.00</td>
<td>-0.60</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Exploratory Factor Analysis

An exploratory factor analysis (EFA) was used because the instrument utilized in this study has only been used in one other study, so the EFA was used to explore and discover the instrument’s properties such as factor structure and utility. An EFA was conducted for four variables using the Kaiser criterion for determining the number of factors to retain. Additionally, the dimensional structure of factors that underlie the set of variables in the model was verified with a Parallel Analysis technique.

Assumptions

Multivariate normality. To assess the assumption of multivariate normality, the squared Mahalanobis distances were calculated for the data and plotted against the quantiles of a Chi-square distribution (DeCarlo, 1997; Field, 2017). In the scatterplot, the solid line represents the theoretical quantiles of a normal distribution. Normality can be assumed if the points form a relatively straight line. The scatterplot for normality was presented in Figure 2.

Factorability. To assess the factorability of the data, Pearson correlations were calculated to determine the intercorrelations for each variable. According to Tabachnick and Fidell (2019), correlation coefficients should exceed .30 in order to justify comprising the data into factors. All variables had at least one correlation coefficient greater than .30 and appear suitable for factor analysis.

Multicollinearity. Although variables should be intercorrelated with one another, variables that are too highly correlated can cause problems in EFA. To assess multicollinearity,
Figure 2. Mahalanobis distance scatterplot testing multivariate normality.

indicates that multicollinearity exists in the data (Field, 2017). The value of the determinant for the correlation matrix was 0.422, indicating that there was no multicollinearity in the data.

Results

The factor loadings were interpreted by taking the absolute value of each loading and implementing the criterion suggested by Comrey and Lee (2013). Values greater than .71 are considered “excellent,” values between .63 and .71 are “very good,” values between .55 and .62 are “good,” values between .45 and .54 are “fair,” and values between .32 and .44 are “poor.” Tabachnick and Fidell (2019) also recommend that .32 should be the minimum threshold used to identify noteworthy factor loadings.
Determination of the number of factors. The Kaiser criterion was chosen for electing how many factors to retain. According to this rule, all factors that have an eigenvalue greater than one are retained for interpretation. The observed eigenvalues were extracted from the correlation matrix with the diagonal of the matrix being replaced by each variable's squared multiple correlations (Ledesma, 2007; Montanelli & Humphereys, 1976) to estimate each variable's communality (DiStefano et al., 2009; Stewart & Ware, 1992). Kaiser's eigenvalue-greater-than-one rule was a simple and common practice used throughout research (Floyd & Widaman, 1995; Ledesma & Valero-Mora, 2007; Yong & Pierce, 2013). Figure 3 shows the scree plot along with the Kaiser criterion for determining the number of significant factors. Looking at Figure 3, there was one factor that had an eigenvalue greater than one.

Figure 3. Scree plot with the Kaiser criterion.
I verified the scree plot/Kaiser 1 (K1) Rule method by conducting a Parallel Analysis to assure that the structure was indeed unidimensional. For this selection method, uncorrelated normal variables were randomly generated that parallel the data in the number of variables and sample size. Next, the observed eigenvalues were extracted from the correlation matrix with the diagonal of the matrix being replaced by each variable's squared multiple correlations (Ledesma, 2007; Montanelli & Humphereys, 1976) to estimate each variable’s communality (DiStefano et al., 2009; Stewart & Ware, 1992). These observed eigenvalues were then compared to the eigenvalues of the randomly generated variables. The Parallel Analysis indicated a mean eigenvalue of .819 and a percentile eigenvalue of .906, so the results from the scree plot were verified. The actual eigenvalues that had a higher value than its randomly generated counterpart were retained for interpretation (Ledesma & Valero-Mora, 2007; Hayton et al., 2004). There were no factors that had a greater eigenvalue than its randomly generated counterpart. As a result, one factor was used for the EFA.

**Evaluating Sample Size.** The sample size for exploratory factor analysis was very important when constructing repeatable and reliable factors. According to Osborne & Costello (2004), the most common guideline for the ratio of sample size to the number of variables (participant to item ratio) included should be at least 10 to 1, but some research indicates a minimum ratio of 5 to 1. The participant to item ratio for this analysis was approximately 33 to 1, where the sample size was 133 and the number of variables included was 4. This indicates that the given sample size was sufficient to produce reliable results.

**Factor summary.** The one-factor model accounted for 39.16% of total variance in the data. The factor analysis summary is shown in Table 9. A Chi-square goodness of fit test was conducted to determine if the one-factor model fit the data sufficiently based on an alpha value
of 0.05, $\chi^2(2) = 14.89, p < .001$. This indicates that the one-factor model did not adequately depict the data, which means that there was a lack of data fit for the model.

### Table 9

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.57</td>
<td>39.16</td>
<td>39.16</td>
</tr>
</tbody>
</table>

*Note: $\chi^2(2) = 14.89, p < .001$.*

**Factor interpretation.** The following variables had excellent loadings for Factor 1: Progressive_Perspective. The following variables had very good loadings for Factor 1: Communication. The following variables had good loadings for Factor 1: Accuracy. Any other loadings were insignificant for Factor 1. The factor analysis loadings are shown in Table 10.

### Table 10

**Factor Loadings From Exploratory Factor Analysis**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor loading</th>
<th>Communality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>0.64</td>
<td>0.41</td>
</tr>
<tr>
<td>Progressive_Perspective</td>
<td>0.88</td>
<td>0.78</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.61</td>
<td>0.37</td>
</tr>
<tr>
<td>AchvVSBeh</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Factor loadings < .32 are suppressed.*
Evaluating the factor structure. According to Costello and Osborne (2005), examining the communality of each variable, checking for crossloadings across multiple factors, and inspecting the number of strong loadings for each factor are good ways to analyze the validity of the factor structure. The following variables had a low communality (< .40): Accuracy and AchvVSBeh. In this case, AchvVSBeh had a communality of 0.00, which indicates that this is a variable with variance that is completely unexplained by any other variables. Each factor had at least three significant loadings (> .32), which was indicative of a strong and solid factor (Osborne & Costello, 2005).

MANOVA

A multivariate analysis of variance (MANOVA) was conducted to assess if there were significant differences in the linear combination of Progressive_Perspective, Communication, AchvVSBeh, and Accuracy between the levels of YearsED_recode, Gender, EdDegree, and Division. I used G*Power apriori planning with alpha .05 and power .8 to determine that a sample size of 100 was needed to be meaningful. The G*Power sample size was exceeded, with a total of 133 participants.

Assumptions

Multivariate normality. To assess the assumption of multivariate normality, the squared Mahalanobis distances were calculated for the model residuals and plotted against the quantiles of a Chi-square distribution (DeCarlo, 1997; Field, 2017). In the scatterplot, the solid line represents the theoretical quantiles of a normal distribution. Multivariate normality can be assumed if the points form a relatively straight line. Strong deviations could indicate that the
parameter estimates are unreliable and multivariate normality cannot be assumed. The scatterplot for normality was presented in Figure 4. Multivariate normality was met in this instance.

Absence of multicollinearity. A correlation matrix was calculated to examine multicollinearity between the dependent variables. All variable combinations had correlations less than 0.9 in absolute value, indicating the results are unlikely to be significantly influenced by multicollinearity. The correlation matrix was presented in Table 11.

![Figure 4](image)

Figure 4. Chi-square Q-Q plot for squared Mahalanobis distances of model residuals to test multivariate normality.
Table 11

Correlations Between Dependent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Progressive_Perspective</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Communication</td>
<td>0.56</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. AchvVSBeh</td>
<td>0.07</td>
<td>0.13</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4. Accuracy</td>
<td>0.54</td>
<td>0.40</td>
<td>-0.20</td>
<td>-</td>
</tr>
</tbody>
</table>

Results

The main effect for YearsED_recode was not statistically significant, $F(8, 236) = 0.77, p = .629$, suggesting the linear combination of Progressive_Perspective, Communication, AchvVSBeh, and Accuracy was similar for each level of YearsED_recode. The main effect for Gender was not statistically significant, $F(8, 236) = 1.47, p = .170$, suggesting the linear combination of Progressive_Perspective, Communication, AchvVSBeh, and Accuracy was similar for each level of Gender. The main effect for EdDegree was not statistically significant, $F(4, 117) = 0.98, p = .419$, suggesting the linear combination of Progressive_Perspective, Communication, AchvVSBeh, and Accuracy was similar for each level of EdDegree. The main effect for Division was statistically significant, $F(24, 480) = 2.44, p < .001, \eta^2_p = 0.11$, suggesting the linear combination of Progressive_Perspective, Communication, AchvVSBeh, and Accuracy was statistically significantly different among the levels of Division. The effect size of 0.11 means that academic subdivision accounts for 11% of the variability in the model. The MANOVA results are presented in Table 12.
Table 12

MANOVA Results for Progressive_Perspective, Communication, AchvVSBeh, and Accuracy by Division

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pillai</th>
<th>$F$</th>
<th>df</th>
<th>Residual df</th>
<th>$p$</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YearsED_recode</td>
<td>0.05</td>
<td>0.77</td>
<td>8</td>
<td>236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.09</td>
<td>1.47</td>
<td>8</td>
<td>236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EdDegree</td>
<td>0.03</td>
<td>0.98</td>
<td>4</td>
<td>117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Division</td>
<td>0.43</td>
<td>2.44</td>
<td>24</td>
<td>480</td>
<td>&lt; .001</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Hypothesis Testing

Research Question 1

Research question one set out to examine what differences exist in community college full-time faculty perceptions of grading. Research question one is as follows:

1. What differences exist in community college full-time faculty perceptions of grading?
   a. How are perceptions different by academic subdivision?
      • Null Hypothesis: H01a: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$
      • Alternative Hypothesis: H11a: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \neq \mu_7$
   b. How are perceptions different between faculty with and without an education degree?
      • Null Hypothesis: H01c: $\mu_1 = \mu_2$
      • Alternative Hypothesis: H11c: $\mu_1 \neq \mu_2$
   c. How are perceptions different by gender?
      • Null Hypothesis: H01c: $\mu_1 = \mu_2$
      • Hypothesis: H11c: $\mu_1 \neq \mu_2$
Through the use of the quantitative survey, as well as an exploratory factor analysis, the research was able to define four latent variables, that created the construct of grading. A MANOVA was then run to answer Research Question 1. Research Question 1 was broken up into three parts, examining perceptions of grading by *academic subdivision*, having or not having an *education degree*, and *gender*. The results of the MANOVA lead to rejecting the null hypothesis for Research Question 1a. The null hypothesis states there will not be differences in perceptions of grading by *academic subdivision*, but the analysis of the MANOVA shows that there is, in fact, a difference among subdivision, as indicated by the \( p \) value of \(< .001\). The effect size (\( \eta^2 \)) was \(.11\), which means that 11% of the variability in the model can be explained by subdivision. As such, Research Question 1a rejects the null hypothesis and accepts the alternative hypothesis, which means that differences exist among full-time faculty members' perception of grading.

As for Research Question 1b, which examined differences in perception between faculty with and without an education degree, results indicated to fail to reject the null hypothesis and there was no mean difference based on having or not having an education degree. Similarly, Research Question 1c, which examined differences in perception based on gender, results indicated to fail to reject the null hypothesis. This means that there were no mean differences in perception of grading based on gender.

**Research Question 2**

Participants were asked what they believe is the purpose of grades, and they were given six options for responses: 1. to communicate the achievement status of students to their parents and others, 2. to document students' performance in order to evaluate the effectiveness of
instructional programs, 3. to communicate with stakeholders, 4. to provide incentives for students to learn, 5. to provide evidence of students' lack of effort or inappropriate responsibility, and 6. to select, identify, or group students for certain educational programs. When looking at the frequency of responses to the question of the purpose of grades, as seen in Figure 5, there are three responses that are most common: to communicate the achievement status of students to their parents and others (1); to document students' performance in order to evaluate the effectiveness of instructional programs (2); and to provide incentives for students to learn (4).

For purposes of this study, it was important to separate responses by having or not having a degree in education to determine if there were differences. The most frequently observed category of Purpose for faculty with no degree in education was to communicate the achievement status of students to their parents and others (n = 26, 32%). Frequencies and percentages are presented in Table 13.
Table 13

Frequency for Faculty without A Degree in Education

<table>
<thead>
<tr>
<th>Variable – Purpose</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>To communicate the achievement status of students to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>their parents and others</td>
<td>26</td>
<td>32.50</td>
</tr>
<tr>
<td>To provide incentives for students to learn</td>
<td>18</td>
<td>22.50</td>
</tr>
<tr>
<td>To communicate with stakeholders</td>
<td>9</td>
<td>11.25</td>
</tr>
<tr>
<td>To document students’ performance in order to evaluate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the effectiveness of instructional programs</td>
<td>21</td>
<td>26.25</td>
</tr>
<tr>
<td>To select, identify, or group students for certain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>educational programs</td>
<td>2</td>
<td>2.50</td>
</tr>
<tr>
<td>To provide evidence of students’ lack of effort or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inappropriate responsibility</td>
<td>1</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Note. Due to rounding errors, percentages may not equal 100%.

The most frequently observed category of Purpose for faculty with a degree in education was to provide incentives for students to learn \(n = 16, 30\%\). Frequencies and percentages are presented in Table 14.

Table 14

Frequency for Faculty with A Degree in Education

<table>
<thead>
<tr>
<th>Variable – Purpose</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>To communicate the achievement status of students to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>their parents and others</td>
<td>14</td>
<td>25.93</td>
</tr>
<tr>
<td>To provide incentives for students to learn</td>
<td>16</td>
<td>29.63</td>
</tr>
<tr>
<td>To communicate with stakeholders</td>
<td>7</td>
<td>12.96</td>
</tr>
<tr>
<td>To document students’ performance in order to evaluate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the effectiveness of instructional programs</td>
<td>14</td>
<td>25.93</td>
</tr>
<tr>
<td>To select, identify, or group students for certain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>educational programs</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>To provide evidence of students’ lack of effort or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inappropriate responsibility</td>
<td>2</td>
<td>3.70</td>
</tr>
</tbody>
</table>

Note. Due to rounding errors, percentages may not equal 100%.
When examining the frequency of purpose selected among those with and without a degree in education, the top three responses are the same among each group. The remaining three responses accounted for less than 16% of the total responses for both groups. To further analyze the data, a one-way analysis of variance (ANOVA) was conducted to determine whether there were mean differences in Purpose by the factors EdDegree, YearsED_recode, Division, and Gender.

**Assumptions**

**Normality.** The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 6 presents a Q-Q scatterplot of model residuals. In this instance, normality was met.

**Outliers.** To identify influential points, Studentized residuals were calculated and the absolute values were plotted against the observation numbers (Field, 2017; Pituch & Stevens, 2015). Studentized residuals are calculated by dividing the model residuals by the estimated residual standard deviation. An observation with a Studentized residual greater than 3.15 in absolute value, the 0.99 quantile of a t distribution with 129 degrees of freedom, was considered to have significant influence on the results of the model. Figure 7 presents the Studentized
Figure 6. Q-Q scatterplot for normality of the residuals for the regression model.

Figure 7. Studentized residuals plot for outlier detection.
Results

The ANOVA was examined based on an alpha value of 0.05. The results from the omnibus ANOVA model indicated that there was, \( F(11, 118) = 2.07, p = .027 \) a statistically significant mean difference in Purpose among at least one of the factors (Table 15). The factor, EdDegree, was not statistically significant, \( F(1, 118) = 0.94, p = .333 \), indicating there were no mean differences in Purpose by EdDegree levels. The factor, YearsED, was statistically significant, \( F(2, 118) = 5.83, p = .004, \eta^2_p = 0.09 \), indicating there were mean differences in Purpose by YearsED levels. The effect size between the demographic, YearsED, and the dependent variable, Purpose, was 0.09. This indicates that 9% of the variability in the Purpose of Grading can be explained by the number of years in education. The factor Division, was not statistically significant, \( F(6, 118) = 1.39, p = .225 \), indicating there were no mean differences in Purpose by Division levels. The factor, Gender, was not statistically significant, \( F(2, 118) = 0.40, p = .671 \), indicating there were no mean differences in Purpose by Gender levels. Figure 8 shows the means differences based on the number of years in education.

Table 15
Analysis of Variance for Purpose by EdDegree, YearsED_recode, Division, and Gender

<table>
<thead>
<tr>
<th>Term</th>
<th>SS</th>
<th>Df</th>
<th>F</th>
<th>p</th>
<th>\eta^2_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>EdDegree</td>
<td>1.49</td>
<td>1</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YearsED_recode</td>
<td>18.41</td>
<td>2</td>
<td>5.83</td>
<td>.004</td>
<td>0.09</td>
</tr>
<tr>
<td>Division</td>
<td>13.15</td>
<td>6</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.27</td>
<td>2</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>186.35</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It was appropriate to conduct the Tukey post-hoc test because there was statistical significance found in the ANOVA. For the statistically significant factor, YearsED_recode, the Tukey post-hoc test was calculated between each pair of measurements to further examine the differences among the variables. Tukey pairwise comparisons were conducted for all combinations of the factor based on an alpha of 0.05. For the main effect of YearsED_recode, the mean of the dependent variable, Purpose, for < 7 years ($M = 1.60$, $SD = 0.74$) was smaller than for 8-15 years ($M = 2.82$, $SD = 1.40$), $p = .003$. The purpose of the Tukey post-hoc test was to determine exactly where differences existed within the factor. This post-hoc test performed pairwise comparisons for the YearsED_recode because the results from the ANOVA indicate

Figure 8. Means of Purpose by YearsED.
that there is statistical significance with that factor. This result can help to explain the answer to Research Question 2 because it gives further insight into which categories of the independent variable are statistically significant. In this case, the comparison of < 7 years in Education with 8-15 years in Education was found to be statistically significant. Furthermore, this finding shows that the purpose of grades does differ among community college full-time faculty. No other statistically significant post-hoc results were found for the levels comprising this factor.

2. How does the purpose of grades differ among community college full-time faculty?

- Null Hypothesis: H02: μ 1 = μ 2 = μ 3 = μ 4 = μ 5 = μ 6
- Alternative Hypothesis: H12: μ 1 ≠ μ 2 ≠ μ 3 ≠ μ 4 ≠ μ 5 ≠ μ 6

Qualitative Data Collection and Analysis

Data Collection Procedures

As described previously, I selected eight individuals with whom to conduct semi-structured interviews. At the end of the survey, participants were invited to indicate their interest in participating in the interview. There were a total of 71 participants who volunteered to be interviewed. Of those 71, eight were from the Nursing and Health Sciences subdivision; 10 were from Science, Technology, Engineering and Math; 17 were from Social and Behavioral Sciences and the Library; eight were from Arts, Culinary, and Hospitality; nine were from Business and Applied Technology; two were from Counseling and Advising; and 17 were from Liberal Arts (Table 16).

Once the initial volunteers were identified, I examined the demographics of those volunteers to ensure that four had a degree in education and four did not. I also chose four
females and four males, and at least one person from each of the different subdivisions. The following table represents the interviewees.

Table 16
Demographics of Semi-Structured Interviewees

<table>
<thead>
<tr>
<th>Name</th>
<th>Subdivision</th>
<th>Gender</th>
<th>Degree in Education</th>
<th>Years in Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sara</td>
<td>Arts, Culinary, and Hospitality</td>
<td>Female</td>
<td>No</td>
<td>8-15 years</td>
</tr>
<tr>
<td>Dave</td>
<td>Business and Applied Technology</td>
<td>Male</td>
<td>Yes</td>
<td>8-15 years</td>
</tr>
<tr>
<td>Jake</td>
<td>Business and Applied Technology</td>
<td>Male</td>
<td>No</td>
<td>8-15 years</td>
</tr>
<tr>
<td>Matt</td>
<td>Counseling and Advising</td>
<td>Male</td>
<td>Yes</td>
<td>&gt; 15 years</td>
</tr>
<tr>
<td>Jill</td>
<td>Liberal Arts</td>
<td>Female</td>
<td>Yes</td>
<td>8-15 years</td>
</tr>
<tr>
<td>Brian</td>
<td>Nursing and Health Sciences</td>
<td>Male</td>
<td>No</td>
<td>&gt; 15 years</td>
</tr>
<tr>
<td>Laura</td>
<td>Science, Technology, Engineering, and Math</td>
<td>Female</td>
<td>Yes</td>
<td>&gt; 15 years</td>
</tr>
<tr>
<td>Olivia</td>
<td>Social and Behavioral Sciences and the Library</td>
<td>Female</td>
<td>No</td>
<td>&gt; 15 years</td>
</tr>
</tbody>
</table>

Each of the interviewees was contacted via email to inquire about times to conduct the interview. Participants were asked to provide me with three dates/times to conduct the interview. Based on those responses, the interviews were scheduled.

Once the interviews were scheduled, the interviewees were sent consent forms, which they all signed and returned prior to the interview. Each of the interviews took place via Zoom. I used Zoom’s automatic transcript feature and also recorded the interviews. At the end of each interview, I accessed the recordings and transcript to refine the transcripts and identify and correct any misspellings and inaccuracies.
After I refined the transcripts, I sent the transcripts to each of the interviewees to ensure accuracy. One of the interviewees responded with four corrections on the transcript. I reviewed the corrections suggested by the interviewee and made the corrections. Once the corrections were made, and resent to the interviewee, they approved. None of the other interviewees had any corrections. One of the interviewees asked to make sure the full transcript would not be published as part of the study. I confirmed that the full transcript would not be published. The additional seven interviewees each approved their transcript.

**Findings from Qualitative Data**

I utilized structural coding strategies to analyze the qualitative data. Precoding was the first step in coding. Utilizing this strategy allowed me to find quotes that were useful in digging deeper into the data. As I read through the transcripts, a number of quotes were identified as being interesting or prominent. I made note of the quotes by highlighting them for future use. I selected quotes if they carried meaning with regards to the research question. The quotes were then used to begin to develop codes. To find these codes, I analyzed the highlighted quotes and found words to represent the quotes, or portions of the quotes that were significant. Once these codes were determined, I then analyzed the codes to create themes.

To begin this process, I first reviewed each of the codes, and then, I looked back at the codes and how the codes fit back into the interviews. I was able to see the clear ties between the codes and the highlighted quotes and key words. As such, it was then important to take those codes and work to create themes. I was able to see how some of the codes were related. For example, the codes of Experience and Being a Parent had similar intended meanings from the interviewees. Experience was mentioned in three of the eight interviews. Being a Parent was
discussed in two of the eight interviews. In the case of both of these codes, the frequency of each was not as significant as other codes, but once combined, had a more significant frequency. Upon reflecting on and further analyzing the quotes and interviews, I found that Colleagues also fit well with the two previously mentioned codes. I found this to be true because in the quotes from the interviewees, the intended meaning for Colleagues was the discussions that led to growth. These codes were then all combined to form the theme of Personal Growth.

A second theme relative to Research Question 3 was Impact of Students. This theme was discussed in five of the eight interviews. In the interviews that had this theme, the discussion from the interviewees was centered around the experiences the faculty member had with their students. These experiences directly relate to how the students helped the faculty member see their perspective on grading through their own eyes.

The final theme I identified was Professional Development. This theme includes the code Professional Development as well as Coursework. The code Professional Development was discussed in all eight of the interviews. The code Coursework was found in three of the eight interviews. In this case, the code is reflected in the theme, and the code still exists as the theme, Professional Development. Based on my professional experience, it made sense that the codes combined to create the overall theme of Professional Development. Quotes specifically related to these themes will be discussed further in this chapter.

**Data Coding for Research Question 3**

I identified three key areas in the semi-structured interviews that would contribute to answering Research Questions 3 and 4. The first area was the answer to the question of how the interviewees learned to grade. The codes are presented in a Wordle to easily identify the
frequency of the themes. The larger the word, the more often the code occurred. The codes found from the transcripts related to this area are found in Figure 10.

These codes were found through an analysis of the quotes as well as of the transcripts in their entirety (Figure 9). I was able to glean insight into the minds of the interviewees by examining the quotes that initially stood out to me. I found that some codes, such as Professional Development and Evolving Practices occurred in the quotes as well as the transcript much more often. The code Being a Parent was only found twice in the analysis, but it was a significant portion of the development of how both of those interviewees learned how to grade, so I felt it was necessary to include them in the coding. Further discussion of their experiences can be found in the subsequent section where I discuss themes.

Figure 9. Codes for how full-time faculty learned to grade.
Theme 1: Professional Development

While reading through the transcripts, there were some key words that began to emerge. One of the first words that was noticed was that interviewees identified some sort of professional development that contributed to how they learned to grade. According to Laura, “During in-service we’ve had those discussions and, to some degree, professional development over the years.” This same sort of response was also stated by Sara, who teaches in the Arts, Culinary, and Hospitality subdivision, when she said that she has done a lot of reading on the subject. The reoccurrence of this code, led to it becoming a theme. This theme was found in six of the eight interviews as something that influenced how they learned to grade.

Theme 2: Personal Growth

Another idea that was mentioned by the interviewees multiple times was how their practices have evolved over time. Six of the eight interviewees remarked that their grading practices were evolving. According to Sara, “I just wasn’t happy with the standard idea of grading, and it didn’t make a lot of sense to me.” Matthew remarked, “The way I formed my philosophies of grading was most organic.” Jake stated, “I was saying, hey, students need to know how to write, they need to know how to communicate, but as I’ve gotten more mature, I feel like the content is what’s more important than the formatting, grammatical, and all that kind of stuff.” All of these quotes led to developing the theme of evolving practices.

Jill discussed her personal experience, and told me about her personal growth journey:

As an English teacher, they don’t really teach you how to teach, so then you go in and you grade. It’s a disaster, right? Because then you grade for grammar, which is so stupid because you can pay an editor to fix your grammar. When I started, I had no idea, and I was overwhelmed. As a new teacher you’re trying to set clear expectations for your class
and you’re also, if you’re in my case, you’re teaching seventh grade in the South Bronx and you’re getting eaten alive. As I’ve had the space and time to not only become clear myself what I feel is important for my students to take away from my class and knowing how students learn, and just understanding more how it’s all so contrived.

This perspective may be a bit unique in that Jill began her teaching career as a middle school teacher, not a higher education faculty member; however, since she was a trained teacher, it seems reasonable to assume that she would have an even better understanding of grading. This does not seem to be the case, as Jill described her struggle with making grade determinations, as indicated in the previous quote.

David also shared a bit about his personal journey of learning how to grade:

I think the biggest thing I’ve learned throughout my various careers and all of that different education and training, really is to understand that there is no systematic single method that says, this is the best metric to determine someone’s aptitude or what they have learned. Individuals are so unique and different, and when you start looking at learning models, visual learners and auditory learners and things of that sort, that there’s no single definitive way to measure that.

This quote also points to the struggle with making the determination about how to grade. Sara said, “I thought about the things that are important to me philosophically and the frustrations I had when I was in college.” The theme of personal growth became apparent as I analyzed the data.

Theme 3: Impact of Students

The third theme found was the impact of students. For full-time faculty members, the experiences with their students, and the lessons learned from their students have shaped their grading practices. Sara explained how her equity lens, as well as her focus on students, has shaped her approach to grading:
I attribute my change in grading philosophy a lot to equity and practices that meet all students’ needs, as opposed to a couple of students who learn in a specific way. I think I’ve been thinking about that just very broadly in the way that I teach and then also just the lack of comfortability I had with assigning numbers. They felt very arbitrary to me and the idea that I’m really interested in growth, and when I assign a number, there’s a finality to it, where the student is like okay, I’m done. Now I can move on the next thing. I’m most interested that they can think about projects as ongoing as opposed to you try it and then you’re done. The finality of a grade destroyed the ability to do that because they’re like oh well, I got this grade, so now I’m done. Now move on to the next thing, which is completely not interesting to me as a teacher, so I just find grading to be counterproductive.

Data Coding for Research Question 4

The next area examined was what the interviewees reported as what the interviewees reported as the components they included when determining student grades. The codes that emerged can be found in Figure 10.

Figure 10. Codes for components of students’ grades.
The codes that emerged are shown in the Wordle as they appear in frequency. The larger words appeared most often in the interviews. Interestingly, none of the words are significantly larger, as with the previous codes, where there were only a handful of codes, which occurred often. With this section of the analysis, there were significantly more codes that emerged. I then used these codes to determine the themes for this section. There are three themes that address research question 4: Assessments, Behaviors, and Achievement.

**Theme 1: Written Assessments**

Laura, who teaches in the Science, Technology, Engineering and Math (STEM) subdivision, commented that “20-30% of each test is going to have some sort of essay component where they have to convince me that they understand the material conceptually.” This was one of many mentions of a writing assessment, hence the development of a theme. Additionally, Jake, who teaches in the Business and Applied Technology subdivision, said, “If they (students) can support their response with the textbook or, especially in business, it’s very content heavy, they’re supporting their writing with the textbook or with solid research. That’s what I want to see.” Jake went on to say that when he was “starting out, coming out of (his) own doctorate, which was so writing heavy,” there was much more of a focus on writing. Jake went on to explain that he has relied on formatting and proper English a bit, but it is still a core component. It is clear that writing is an important component currently in full-time faculty grades at TSA.
Another theme found was achievement. This was a theme that was found in half of the interviews. Matt, who teaches in the Counseling and Advising subdivision, remarked,  

I wanted the grade to show growth. I think the classes I felt best about were the ones when I felt at the end of the term, I knew more than at the beginning, so I tried to weight my assignments to be so they had to genuinely grow within themselves to earn that A.

Additionally, Jill, who teaches in the Liberal Arts subdivision, described how she grades by saying, “Are they (students) moving towards the expectations? We have those conversations as a class; however, I don’t grade that. If they’ve done it, and they’re showing growth, then they get their full points.” This quote epitomizes the conflict that grading creates between a faculty member knowing what they want students to be able to do and a faculty member being able to document that growth in a grade.

According to Olivia, who teaches in the Social and Behavioral Sciences subdivision:

When you start to see students who are in your presence and they’re participating and they love the materials, but that does not meet out on their exam, you start to recognize there are some other way that you could be doing this. I’m not really sure what that is, but there has to be a better way for a student who doesn’t show up to class and comes in, maybe once between now and the test, and is able to come in on the review session, get the information, and score really high. And then you have the student who’s there every single class, participating. You really feel like they’re getting it, and their test grade doesn’t reflect that. I think that’s when you start to say I need to put some other things in place to really capture this student’s difficulty with test taking, or they just need another avenue to showcase their abilities and the things they’re learning.

Olivia went on to say:

I think what I’ve always done, I think what you see other educators do. Sometimes you try things and it’s a lot more work, and people start to tell you, why are you doing so much? You can just throw a quiz out there and throw a test out there. And if you’re not careful, I think you will do that.
By examining these quotes, it becomes clear that there is internal conflict regarding what should be included in students’ grades. I found this type of conflict in seven of the eight the interviews. It seemed as though the interviewees felt they had to defend what they decided to include in their grades. Based on the interviews, there is definitely a lack of confidence in how students’ grades are calculated and reported. This idea is further elaborated on by Jill:

This writing’s a bunch of bullshit, grading a bunch of bullshit, but give feedback on growth. It’s all a contrived system. You should be grading towards if students are meeting the standards. I have played around with a lot of different grading. I tried a semester of self-grading, of not giving grades, and now the way I grade, as I start the semester, I say my expectation is that everyone is going to get an A.

The level of internal conflict can be heard in this quote from Sara, who teaches in the Art, Culinary, and Hospitality subdivision:

Students say they want you to give them a quiz and grade it, but I ask, why? I can give them a quiz and grade it, and if you do well, it counts towards your grade, and if your grandma died and you haven’t been in class for two weeks and we had an in-class quiz, I’m just not going to put it in the gradebook.

The internal conflict continues to be apparent in this quote from Jill, “I’m asking myself if I prepared them to go on, for the demands of college and reading and writing and have they met the objectives of the course, and there’s a lot of subjectivity.”

According to Olivia,

Assignments are structured around the course objectives, and what I know from my experience of over 20 years as a teacher, about what they need to be successful, and they leave my class, and they have to read and write in their Biology class. So that’s what I grade on.

Unpacking this quote led me to identify that full-time faculty struggle to find ways to show that grades match what they actually want to report on for their students.

David discussed how grading assessments impacts student learning:
I don’t know that the grades are always as reflective of student learning. I don’t necessarily believe that to be broadly true. I have a significant number of students that simply have testing anxiety, that if I tell them the final exam is a multiple-choice exam, that will be proctored through the testing center, those students do very poorly because of the testing model that we use in this environment. I actually have two final exams for that purpose, there’s a written, multiple choice, but then what I call a practical, where the students now execute and build the skill sets that they’ve learned, and very often I’ll find a student who does poorly on a multiple-choice written exam, but when it comes to that practical, they are head and shoulders above their peers. So now, the question becomes, what skill set is most beneficial to a student who’s progressing into a career field like IT.

**Theme 3: Behaviors**

In three of the interviews, the interviewees discussed how they weighed behaviors in their grades. Behaviors include criteria such as attendance, tardiness, and turning work in on time. Sara described her decision to eliminate attendance from her grades:

A couple of years ago I started no longer grading attendance, as kind of a test point, to see if people showed up because they got points for attendance, or if they showed up here engaged with the material. I felt very nervous about it because whenever you take away structure, what does that mean? So I started with a carrot, if you attend every single class, you get these extra points. Then, I discovered that the students really were on. It was more the policies just made for difficult situations that would happen, regardless of whether or not I had those policies, and it gave me a lot less flexibility in terms of meeting the needs of individual students on their learning journeys.

Engaging in discussions with students regarding how they believe their grades should be calculated was a process Jake engages in with his students. Here, Jake, described his process of engaging with students to define grades:

I ask, how do you think I can measure your capability of doing this? And we will collectively come up with different methods and different processes. What I have historically learned now is that the majority of students don’t see a lot of benefit to quizzes, pop quizzes. From their learning perspective, some will say it puts pressure on them to read and study material, but it doesn’t necessarily mean that they’re absorbing content, and they will apply it in a practical skill format, so I think that’s another big factor in using students to provide that feedback as to how do we measure their education, their success.
Interestingly, the students did not put an emphasis on behaviors for the class, but they did focus on feedback. He went on to say:

I’m very frustrated with higher education because I feel like when I teach a class, students are more worried about the A, B, or C than they are what I’m saying. So I’ve often debated about what’s the best way they’re really going to digest what I’m putting out, and maybe not be as concerned on the A, B, C or D. However, we are almost all held to that standard, to a certain extent, so I have to walk away. I have to have them walk away with some sort of measurement. All we have right now is that grade, that GPA.

Jake’s remarks regarding what he wants to be able to do in the classroom versus what he is expected to do speaks to the cultural expectations with regards to grading. This frustration was also seen in four interviews when instructors spoke of wanting to eliminate behaviors from their grading but felt unable to do so due to expectations of their students as well as the administration.

Further Analysis

The third area I analyzed was what full-time faculty members would include in students’ grade in an ideal world. The codes can be found in Figure 11.

Looking at the Wordle, it is clear to see that one code, application, emerged most frequently. According to Jill, “It’s not so much I want students to memorize things and be able to regurgitate things as I’m interested in them being able to process things so that if I’m there or whomever is there, they can make their own decisions.” This quote highlights to desire for the students to be able to apply what they are learning.

I believe that although the faculty members may have used different terms, they might all be discussing similar topics when they talk about mastery, effort, and growth. Further examination of these would prove to be beneficial to the field.
In examining Research Question 3, what criteria do community college full-time faculty use to determine student grades, the answer to that is difficult. The aforementioned themes are all components that full-time faculty members report to be part of students’ grades; however, there was little consistency among the different faculty members. Reflecting on Figure 11, it becomes logical that there are some components that may be found more often in grades among full-time faculty members, but how those grades are weighted or how they are measured varies greatly. The components of grades are: growth, completion grades, activities, projects, midterms/final, assessments, journals, current events, discussions, attendance, papers, reflections, effort, participation, and writing.
Having a better understanding of what faculty members include as part of students’ grades can help the field to move forward. In this case, there is a wide variety of components, but those components can all be filtered into the themes: Assessments, Achievement, and Behaviors. The wide variety of components highlights the inconsistencies how the full-time faculty members determined students’ grades. From a student perspective, this can cause frustration because each faculty member has such different grading components. Given this understanding, it is logical that students who have learned how to adapt to different grading criteria receive better grades. This idea is also referred to as playing the game of school. For those students who do not have the skillset to effectively navigate how different instructors structure their grading, finding success could be difficult. Essentially, grading can be a barrier to student success if students do not play the game of school well.

Research Question 4

Research Question 4 asks what past experiences do community college full-time faculty attribute to their learning how to determine grades? Answering this question succinctly would be extremely difficult but examining the Wordle in Figure 10 truly provides insight into the minds of full-time faculty and how they came to learn how to grade. The themes found to answer this research question are: professional development, personal growth, and impact of students.

Examining these themes, I can see that full-time faculty members have learned how to grade not only through professional development, but also from their personal experiences as well as from their students. Similar to the wide variety of components of grading, how faculty members learned to grade also stems from a wide variety of experiences. In some cases, faculty members were intentional about seeking out opportunities to increase their capacity for grading,
which led to changes in their grading practices. In other cases, faculty members had students, whose experiences in their classes, led them to investigate their grading practices. From this, I recognized the fluidity of grading. A students’ grade in a course could change in any given semester, based on the experiences of the faculty member. Due to this fluidity, there could be significant changes in how faculty members grade from semester to semester.

I feel it is also important to recognize how different faculty members’ experiences in school shape their grading practices. In some cases, faculty members continue to grade how they were graded because they were fond of the grading system. In these cases, the faculty member was well-versed in playing the game of school, which further perpetuates playing the game. In other cases, the faculty member was not satisfied with their own personal experiences in school, and they sought out additional opportunities for personal growth regarding their own grading practices. In yet other cases, the faculty member had instances where the experience of the students in their class impacted how they shaped their grades. Currently, the field of education lacks research on how to best support student success through reporting student learning, so guidelines for doing so, especially in higher education, are lacking. As with the findings from Research Question 3, there are inconsistencies with how faculty members learned how to grade, which means there are inconsistencies in how faculty members are grading in their own classroom.

Conclusion

Analyzing quantitative and qualitative data allowed me to answer the four research questions of the study. The use of inferential statistics led to rejecting the null hypothesis for
research questions 1 and 2. Analysis of qualitative data led to the determination of themes to answer research questions 3 and 4.
CHAPTER 5
DISCUSSION AND IMPLICATIONS

Discussion of Major Findings

For decades, higher education has been relying on an imperfect system to report student learning. This study aimed to answer four research questions, all which addressed grading in some respect. Although the field of education is lacking in research on grading in higher education, I was able to find research from K-12 that was applicable to higher education, as well as a variety of studies that have been done recently that relate to the topic. The independent variables for this study were Gender, Years in Education, Degree in Education, and Academic Subdivision. These variables were derived from the literature and guided this study.

Impact of Academic Subdivision

One of the major quantitative findings of this research study was that among full-time faculty members, the academic subdivision to which the faculty member belongs had an impact on their grading. For Research Question 1, which asked what differences exist in community college full-time faculty perceptions of grading, I rejected the null hypothesis that said no means differences existed, and accepted the alternative hypothesis, which said that means differences did exist. The acceptance of the alternative hypothesis for the research question opens the field to more research because it highlights how one demographic, in this case academic subdivision, can cause a difference in how a faculty member determines their grades for a course. Based on this finding, grading should be looked at critically through the lens of academic subdivision, and
faculty members should be called on to recognize that the determination student grades is
dependent on the academic subdivision in which the faculty member resides. This finding has
significant implications for practice moving forward because the finding highlights the
inconsistency in grading by academic subdivision at the community college, and it calls into
question the subjectivity of grading.

This information is important to the field because it helps to identify areas in which
further research should be done. Research Question 1 in this study found that the subdivision to
which the faculty member belongs has a significant impact on how they grade. The literature
suggests that grading varies greatly, and the answer to Research Question 1 in this study begins
to suggest a root of the issue, that grading varies by academic subdivision (Cizek et al., 1996;
Guskey, 2009; McMillan et al., 2002). The statistically significant difference among faculty
members by academic subdivision describes the divide in the perception of grades based on
which subdivision the faculty member resides. Further research would need to be done to
determine which subdivisions are statistically significant from one another and to determine how
to reconcile this issue. A fundamental question would be to ask if it is an equitable practice that
grading varies by academic subdivision. This result is significant because it leads to even more
questions, such as why is there a difference in the perception of grading based on academic
subdivision?

**Impact of Years in Education**

Research Question 2, which asked how the purpose of grades differ among community
college full-time faculty, was also examined using quantitative data. Based on the results of the
ANOVA, I rejected the null hypothesis, which said there were no means differences in the
purpose of grades among community college full-time faculty and accepted the alternative hypothesis which said that there were means differences in the purpose of grades by way of the independent variables, *Gender, Academic Subdivision, Years in Education, and Degree in Education*.

This study showed that the purpose of grades varies significantly based on the years the faculty member has been in the field of education. Coupling this quantitative data with the qualitative data, it becomes clear that the statistical answer to Research Question 2, that the purpose of grades varies by the number of years in the field, was corroborated with the qualitative findings and themes of Professional Development, Personal Growth, and Impact of Students. It is logical that with more years of experience, the faculty member has more professional development and growth, and that more students have had an impact on the faculty member. While the literature suggested that having the number of years in education as a variable, the findings in this study suggest that more research must be done to examine what changes in grading come with years of experience.

Furthermore, one of the themes that was found to answer the Research Question 3, which asked what criteria faculty members use to determine student grades, was Personal Growth. The ideas that the interviewees discussed related to personal growth include the evolution of their grading practices and experience in the field. One of the areas that was highlighted by faculty members under the theme of personal growth was evolving practices. Thus, even faculty members who have been in the field for quite some time experience an evolution in the way they grade. Furthermore, the grades that students are receiving in class today, may not be the same representation of student learning for students who took the same class even just last semester. These findings align with the findings for Research Question 2 that indicate that the number of
years in education makes a difference in the faculty member’s belief in the purpose of grading. This information is significant because it highlights the value of having spent years as a faculty member. If there is an institutional focus on grading, seeking out faculty members with more years of experience in the field would be beneficial.

**Impact of Learning to Grade**

An additional finding of this research study is that there is no clear way that faculty members learned how to grade. The analysis of the qualitative data gathered for Research Question 3 highlighted the inconsistencies in how grades are determined at TSA community college, an issue that stems from the fact that there is no consistent direction regarding the best way to formulate grades in higher education.

I found there were many different components included in students’ grades. In fact, it was difficult to capture all of the different items because there was such a wide variety of them. Interestingly, the items that were repeated most often fell under the theme of Assessments. With all of the focus on assessment, this finding means that faculty members are trying to find ways to appropriately grade their students, but that they are struggling to do so. In progressive grading methods, behaviors are reported separately, so this finding could indicate that faculty members are trying to move in that direction, although further research would be needed to make that determination. It would also be important for the faculty members to identify what goes into the grade on an assessment in order to fully determine if behaviors are being reported separately.

While some of the faculty members based their grading on growth, which is an extremely subjective practice, others based it on assessments or even effort, which is even more subjective. In the interviews, there were also numerous behaviors that were reported to be part of students’
grades, such as attendance, completion grades, and participation. The variety of components in grades in higher education, compounded with faculty members’ variety of experiences and lack of training on how to grade all signal a significant area of improvement. At this point, I conclude that grading is quite subjective and dependent on criteria that the faculty member, or in some cases, the students, decide is important to include in a grade.

Implications for Practice

This research study begins a conversation about grading in higher education. Historically, there have been few research studies done on this specific topic, especially in higher education. Since the focus in higher education is on faculty being content experts, and faculty members are required to have a master’s degree or higher in their content area, there is often little emphasis placed on the art and science of teaching. This is especially problematic when you take the results of this study, which show that perceptions of grading vary by the academic subdivision in which the faculty member teaches and considering how students take courses from a variety of academic subdivisions at the community college. These inconsistencies can have a significant impact on student success for those students who do not navigate how different instructors structure their grades. These inconsistencies can be a barrier to student success at the community college.

When reflecting on the findings of this study, which show that the purpose of grading varies by years of experience and by the instructors’ experiences personally and professionally, the notion that grading is meaningful becomes clouded with doubt, despite how engrained grading is in our culture. There is so much variation regarding what goes into a grade, what faculty members know about grading, and how they learned to grade, that consistency is nearly
nonexistent. In order to create more meaningful and consistent grades, faculty members would need to be consistent in their grading. However, it is important to note that applying poor grading practices consistently is no better than inconsistent grading practices. Having said that, the best course of action would be to research the most effective grading practices in higher education and then utilizing professional development to train faculty on better grading practices.

In this research study, all of the full-time faculty members who were interviewed openly discussed the evolution of their grading practices, and they identified personal experiences about their own education and about their professional experiences in the classroom as a faculty member, that had an impact on their grading practices. Given this information, there is a significant opportunity to provide a structure for grading that would incorporate those components that are most important to higher education, which could mean letting go of traditional grading practices and moving to more progressive practices.

On the flipside of that implication, this research study also highlights how grading practices are highly individualized and vary greatly from faculty member to faculty member. Another implication of this research study is that the emphasis placed on grading in higher education may be misguided. It is logical to consider what role grading plays in higher education and what role it should play in higher education in the future. It seems clear that finding an adequate way to report student learning should be a priority for higher education. Given the findings in this study, it would be my recommendation that community colleges seek resources on progressive grading practices and determine how to best implement those practices. This means that community colleges should reflect on current grading practices and decide which practices support and which hinder student success. In order to report student learning in an
effective manner, making the determinations mentioned above and then using professional
development to implement those changes is necessary.

Limitations of this Research

Due to the sample for this research, the results are not generalizable. I used a convenience
sample at my place of employment, and although it is a rather large sample, it is not
generalizable. If I had used a sample of full-time faculty from a variety of community colleges, it
would be possible to generalize the results.

An additional limitation of this research study is the lack of prior research on the topic.
As a result of the void of research on this topic, the depth of the literature is lacking. Once
additional research is done to better understand grading at the community college, there will be
more to guide future research studies. I am optimistic that the higher education community is
ready to embrace the topic and conduct further research.

A third limitation to this study was that there were no volunteers for the semi-structured
interviews who had less than seven years of teaching experience. Since the quantitative data
found that the number of years in the field of education is statistically significant when looking at
the purpose of grades, it would have been relevant to interview a faculty member with less than
seven years of experience in the field. However, given the volunteer pool, interviewing someone
with less than seven years of experience in the field proved to be impossible.

Opportunities for Further Research

An opportunity for additional research would be to further examine the purpose of
grading among faculty members in higher education to determine what they, collectively, believe
is the purpose for grades. This study was limited to full-time faculty, but I recommend conducting the research and including adjunct faculty, as the inferential statistics could still be conducted, but just sorted by full-time faculty and adjunct faculty. Although this study provides some statistics for how often respondents selected each purpose, it is unknown as to why they made those selections, or what influenced their responses. Using quantitative research to follow up on statistics from this study would prove to be beneficial to the field. A potential research question would be: What do faculty identify as the purpose of grades in higher education?

Additionally, there is an opportunity to replicate this research across a region, so the results of the study could be generalized. The survey instrument has been utilized for two research studies, and the items in the survey have been verified through statistical analyses, so conducting future research also utilizing the same survey instrument would be beneficial. As suggested above, I recommend using both full-time faculty as well as adjunct faculty for the population in future studies. The research questions for the study could potentially be the same as in this study, since the concepts are the same.

It would also be beneficial to conduct research on effective ways to report student learning from a faculty perspective, as well as from a student perspective. As noted previously, there is a significant internal conflict that exists when determining grades, so having a better understanding of how grading practices impact student success, and how faculty members can better communicate student progress in a course, would be extremely beneficial to faculty members and students alike. Establishing clear guidelines, or parameters, for reporting student learning at the community college, would be beneficial to the field. Now that the research has been conducted and, according to the results from Research Question 4, it is clear that there is no consistent way that full-time faculty members learn how to grade, and that, according to
Research Question 3, there is not consistency among full-time faculty members regarding the components of a student’s grade, research needs to be done regarding the best way to report student learning. A potential research question for this study would ask what components have the most impact on student success in a student’s grade. Furthermore, it would be beneficial to study the impact of grading on students at the community college, especially those students who struggle to navigate how different instructors structure their grades. Assigning grades, especially ones that are inconsistent at best, as found in this research study, could be more of a barrier to student success than something that promotes it.

Conclusion

This research study set out to gain more insight into the perception of grading by community college, full-time faculty members. When I began reviewing the literature to prepare for the study, an initial lack of literature was noted. While this still remains an issue, this study provided an initial understanding of how full-time faculty at the community college perceive grading.

This research study found that there are differences in perceptions of grading based on the academic subdivision to which the faculty member belongs, as discussed in Research Question 1. Additionally, it was determined that there are differences about what full-time faculty members believe is the purpose of grades, as evidenced in the discussion of Research Question 2.

While it was interesting to hear about the variety of components that go into a student’s grade during the interviews, the variety should also be cause for concern because these are inconsistent and different grades could potentially mean different things to different faculty
members and students. Further studies asking to better understand effective grading practices and how to implement them would benefit the field of higher education. This research study is just the beginning of the journey to develop better grading practices that will truly enhance student learning and ultimately lead to student success.
REFERENCES


Wormeli, R. (2013). *The collected writings (so far) of Rick Wormeli.* Association for Middle Level Education.

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL
February 3, 2021

April Zawlocki  
2393 Iroquois  
Ln.  
Yorkville, IL 60560

**Project Name:** Community College Full-time Faculty Members’ Perceptions of Course Grading  
**Principal Investigator(s):** April Zawlocki  
**Starting Data of Project:** 2/3/2021  
**Ending Data of Project:** 2/3/2022  
**IRB Review Type:** Expedited  
**IRB Review Action:** Approved

Dear Ms. Zawlocki:

I am pleased to inform you that after review of your proposal, *Community College Full-time Faculty Members’ Perceptions of Course Grading*, the College of DuPage Institutional Review Board (IRB) has approved the above-referenced submission by Expedited Review.

Please note that the approval of this protocol will lapse on 2/3/2022, at which time you may request an extension.

Any changes in the procedures affecting interaction with human subjects should be reported to the College IRB. Significant changes will require the submission of a revised request for research.

We wish you well in your project, and should you have any questions, please contact me at 217-652-4724 or at bentej@cod.edu.

Sincerely,

Jim Benté
James R. Benté, Vice President
Planning and Institutional
Effectiveness
Co-chair, College of DuPage Institutional Review Board

cc: Dr. Ken Gray, Professor, Psychology & Co-chair, College of DuPage Institutional Review Board Kathy Cosentino, Administrative Assistant VI
APPENDIX B

SURVEY PARTICIPANT CONSENT
Survey Participant Consent

Community College Full-Time Faculty Members' Perceptions of Course Grading

Study Title: Community College Full-Time Faculty Members' Perceptions of Course Grading

Investigators
Name: April Zawlocki Dept: TLCI Phone: 630-802-7175

Key Information
• This was a voluntary research study on grading practices at the community college.
• This portion of a mixed methods study only involves completing a 10-minute online survey.
• The benefits include being able to voice your perceptions regarding your grading practices; there are no risks involved in the participation of this study.

Description of the Study
The purpose of the study was to examine community college full-time faculty members’ perceptions of grading and reporting student learning via course grades. The researcher will be looking to see what differences exist in perceptions of grading, how the purpose of grades differs among faculty members, the criteria faculty members use to determine grades, and what past experiences faculty members attribute to learning how to grade. If you agree to be in this study, you will be asked to do the following things: to complete a one-time, 10-minute survey via Qualtrics.

Risks and Benefits
There are no reasonably foreseeable risks to participating in this study.

The benefit of participation will be the knowledge gained from the research.

Confidentiality
• This study was confidential. We will not be collecting or retaining any information about your identity.
• The records of this study will be kept strictly confidential. Research records will be kept in a password-protected, one drive file, and all electronic information will be coded and secured using a password protected file. The video recordings of the zoom interviews will also be kept in a locked file, and only the researcher will have access to them, solely for the purpose of this study. The files will be erased after four years. I will not include any information in any report I may publish that would make it possible to identify you.

Your Rights
The decision to participate in this study was entirely up to you. You may refuse to take part in the study at any time. Your decision will not result in any loss of benefits to which you are otherwise entitled. You have the right to skip any question or research activity, as well as to withdraw completely from participation at any point during the process.
You have the right to ask questions about this research study and to have those questions answered before, during, or after the research. If you have any further questions about the study, at any time feel free to contact the researcher, April Zawlocki at zawlockia@cod.edu or by telephone at 630-802-7175. You may also contact the researcher’s Dissertation Chair, Dr. Laurie Elish-Piper at laurieep@niu.edu. If you have any questions about your rights as a research participant that have not been answered by the investigators or if you have any problems or concerns that occur as a result of your participation, you may contact the Office of Research Compliance, Integrity, and Safety at (815)753-8588.

By clicking the “>>” button below I am acknowledging that I am a full-time faculty member at the College of DuPage and give my consent to participate in this research study.
Interview Participant Consent

Study Title: Community College Full-Time Faculty Members’ Perceptions of Course Grading

Investigators
Name: April Zawlocki Dept: TLCI Phone: 630-802-7175

Key Information
- This was a voluntary research study on grading practices at the community college.
- This mixed methods study involves an online survey and subsequent interviews with select faculty members.
- The benefits include being able to voice regarding your grading practices; there are no risks involved in the participation of this study.

Description of the Study
The purpose of the study was [explain research question and purpose in lay language]. If you agree to be in this study, you will be asked to do the following things: to complete a one-time, 10 minute survey via Qualtrics. You may also be asked to participate in a 20-minute interview via Zoom.

Risks and Benefits
There are no reasonably foreseeable risks to participating in this study.

The benefit of participation will be the knowledge gained from the research.

Confidentiality
- This study was anonymous. We will not be collecting or retaining any information about your identity.
- The records of this study will be kept strictly confidential. Research records will be kept in a locked file, and all electronic information will be coded and secured using a password protected file. The video recordings of the zoom interviews will also be kept in a locked file, and only the researcher will have access to them, solely for the purpose of this study. The files will be erased after four years. We will not include any information in any report we may publish that would make it possible to identify you.

Your Rights
The decision to participate in this study was entirely up to you. You may refuse to take part in the study at any time. Your decision will not result in any loss of benefits to which you are otherwise entitled. You have the right to skip any question or research activity, as well as to withdraw completely from participation at any point during the process.

You have the right to ask questions about this research study and to have those questions answered before, during, or after the research. If you have any further questions about the study, at any time
feel free to contact the researcher, April Zawlocki at zawlockia@cod.edu or by telephone at 630-802-7175. You may also contact the researcher’s Dissertation Chair, Dr. Laurie Elish-Piper at laurieep@niu.edu. If you have any questions about your rights as a research participant that have not been answered by the investigators or if you have any problems or concerns that occur as a result of your participation, you may contact the Office of Research Compliance, Integrity, and Safety at (815)753-8588.

Your signature below indicates that you have decided to volunteer as a research participant for this study, and that you have read and understood the information provided above. You will be given a signed and dated copy of this form to keep, along with any other printed materials deemed necessary by the study investigators.

________________________________________________  _____________________
Participant’s Signature                                         Date

I give my consent to be recorded via zoom during the individual interview.

________________________________________________  _____________________
Participant’s Signature                                         Date
### Survey Instrument

<table>
<thead>
<tr>
<th>Item</th>
<th>Label</th>
<th>Likert-scale Questions (Scale: “Strongly Disagree,” “Disagree,” “Agree,” “Strongly Agree”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tch w/o grds</td>
<td>Teachers can teach without grades.</td>
</tr>
<tr>
<td>2</td>
<td>Lrn w/o grds</td>
<td>Students can and do learn without grades.</td>
</tr>
<tr>
<td>3</td>
<td>Math prec</td>
<td>Mathematical precision does not yield fairer or more objective grading.</td>
</tr>
<tr>
<td>4</td>
<td>Average</td>
<td>Averaging to obtain a course grade was a questionable practice.</td>
</tr>
<tr>
<td>5</td>
<td>Zeros</td>
<td>Giving zeroes for work missed or work turned in late was a questionable practice.</td>
</tr>
<tr>
<td>6</td>
<td>Infractions</td>
<td>Taking credit away from students for infractions was a questionable practice.</td>
</tr>
<tr>
<td>7</td>
<td>Recent</td>
<td>I give priority to the most recent evidence when determining grades.</td>
</tr>
<tr>
<td>8</td>
<td>Comprehensive</td>
<td>I give priority to the most comprehensive evidence when determining grades.</td>
</tr>
<tr>
<td>9</td>
<td>Import goals</td>
<td>I give priority to the evidence related to the most important learning goals or standards when determining grades.</td>
</tr>
<tr>
<td>10</td>
<td>Beh separate</td>
<td>I report behavioral aspects separately.</td>
</tr>
<tr>
<td>11</td>
<td>Scales</td>
<td>I use grading scales instead of percentages.</td>
</tr>
<tr>
<td>12</td>
<td>Reward</td>
<td>Grades have some value as rewards but no value as punishments.</td>
</tr>
<tr>
<td>13</td>
<td>Criteria</td>
<td>Grading and reporting should always be done in reference to learning criteria-never on the curve.</td>
</tr>
<tr>
<td>14</td>
<td>Percent</td>
<td>High percentages are not the same as high standards.</td>
</tr>
<tr>
<td>15</td>
<td>Stkhdr lang</td>
<td>I use language clearly understood by stakeholders.</td>
</tr>
<tr>
<td>16</td>
<td>Info source</td>
<td>Assessments must be sources of information for student and faculty members.</td>
</tr>
<tr>
<td>17</td>
<td>Correctives</td>
<td>When I give an assessment I give corrective feedback.</td>
</tr>
<tr>
<td>Item</td>
<td>Label</td>
<td>Demographic Questions</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>18</td>
<td>2nd chance</td>
<td>When I give assessments I give students a second chance to show improvement.</td>
</tr>
<tr>
<td>19</td>
<td>Collaborate</td>
<td>I work with other faculty members on classroom assessments.</td>
</tr>
<tr>
<td>20</td>
<td>Explicit</td>
<td>I match assessments to what I am explicitly teaching (i.e. I don’t use trick questions, new formats, or unfamiliar material).</td>
</tr>
<tr>
<td>21</td>
<td>Studs know</td>
<td>My students know the criteria/objective to be assessed.</td>
</tr>
<tr>
<td>22</td>
<td>Oppt for risk</td>
<td>I give students the opportunity to take risks in my classroom.</td>
</tr>
<tr>
<td>23</td>
<td>Hwk&gt;ach</td>
<td>I believe homework increases student achievement.</td>
</tr>
<tr>
<td>24</td>
<td>Stdy grp</td>
<td>I believe that study groups improve student achievement.</td>
</tr>
<tr>
<td>25</td>
<td>Same hwk</td>
<td>All students in my classroom are given the same homework assignments.</td>
</tr>
<tr>
<td>26</td>
<td>Grade hwk</td>
<td>I grade all homework.</td>
</tr>
<tr>
<td>27</td>
<td>Mean hwk</td>
<td>I believe the extent to which my homework was completed relies on how meaningful it was to students.</td>
</tr>
<tr>
<td>28</td>
<td>Consistency</td>
<td>Please answer ‘Always’, ‘Sometimes’, or ‘Never’ to the following statements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I believe the grades I give are consistent:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Within a class</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• From class to class</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• From semester to semester</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• From year to year</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• From instructor to instructor of the same course</td>
</tr>
<tr>
<td>29</td>
<td>Purpose</td>
<td>If the College were to identify the major purpose of grades for report cards, which of the following do you believe that purpose should be:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• to communicate the achievement status of students to their parents and others</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• to document students’ performance in order to evaluate the effectiveness of instructional programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• to communication with stakeholders</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• to provide incentives for students to learn</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• to provide evidence of students’ lack of effort or inappropriate responsibility to select, identify, or group students for certain educational programs</td>
</tr>
<tr>
<td>Item</td>
<td>Label</td>
<td>Demographic Questions</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>30</td>
<td>Gender</td>
<td>Please select your gender:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Male</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Female</td>
</tr>
<tr>
<td>31</td>
<td>Bkgrnd</td>
<td>Please select the option that best describes your educational background:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Degree in an Education field (i.e. Early childhood education, Elementary education, Secondary education, Special Education, Physical Education, Music Education, Curriculum &amp; Instruction)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- No degree in an education field</td>
</tr>
<tr>
<td>32</td>
<td>Sub Div</td>
<td>Please select your academic subdivision at TSA:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Arts, Culinary, and Hospitality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Business and Applied Technology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Liberal Arts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Nursing and Health Sciences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Social and Behavioral Science and the Library</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Science, Technology, Engineering and Math</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Counseling and Advising</td>
</tr>
<tr>
<td>33</td>
<td>Prof Dev</td>
<td>Please describe the amount of professional development you have had to enhance your grading practices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Very little</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Little</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Much</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Very much</td>
</tr>
<tr>
<td>34</td>
<td>Tchg Exp</td>
<td>Please identify the number of years experience you have teaching at the college level:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- less than 3 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 3 to 7 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 8 to 15 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- more than 15 years</td>
</tr>
<tr>
<td>35</td>
<td>Int</td>
<td>Please indicate if you are willing to participate in a Zoom interview with the researcher to expand on your responses.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- No</td>
</tr>
</tbody>
</table>

Guskey (2009)
APPENDIX E

SUPPORT TO REPLICATE STUDY
Support to Replicate Study

Replicating Study

From: April Zawlocki <april.zawlocki@gmail.com>
To: Thomas

Good Evening,

I am a doctoral student at Northern Illinois University, and I am interested in replicating your study from 2009: Bound by Tradition: Teachers’ Views of Crucial Grading and Reporting Issues. The methodology on my dissertation committee suggested that I contact you to receive your permission to replicate the study using your survey. Could you please alert me to any costs that may be involved with using the tool?

Thank you for your time and consideration.

Best regards,
April Zawlocki

From: Thomas Guskey <guskey@uky.edu>
To: me

Hi April,

You are welcome to use the survey. All that I ask is that you reference the original article and inform me of your results so that I may cite your research.

Thanks for your consideration.

Best wishes,

Tom

-------------------------------
Thomas R. Guskey, Professor
College of Education
University of Kentucky
Lexington, KY 40506

Ph: 859-221-0077
Email: guskey@uky.edu
Twitter: @tguskey
Web: http://education.uky.edu/EDP/guskey
APPENDIX F

SEMI-STRUCTURED INTERVIEW PROTOCOL
Semi-Structured Interview Protocol

1. What training, if any, have you received on educational measurement (grading)?
   a. Where did the training take place?
   b. When did the training take place?
   c. What did you learn from the experience?
   d. Did you learn anything that affects your grading practices? If so, what?
2. What are criteria you use to determine student grades?
3. How did you arrive at those criteria?
4. Have you always used the same criteria?
   a. If not, how have the criteria changed over time?
   b. Why have there been changes?
5. What has influenced your decision about the criteria?
6. If you were to create the ideal situation, what criteria would determine a student’s grade?
   a. Why?