

2022

Examining the Ways Female Students Experience Teacher Feedback and Assessment

Nicole Foran

Follow this and additional works at: <https://pilotscholars.up.edu/etd>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Gender Equity in Education Commons](#), and the [Secondary Education Commons](#)

Recommended Citation

Foran, Nicole, "Examining the Ways Female Students Experience Teacher Feedback and Assessment" (2022). *Graduate Theses and Dissertations*. 106.
<https://pilotscholars.up.edu/etd/106>

This Doctoral Dissertation is brought to you for free and open access by Pilot Scholars. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Pilot Scholars. For more information, please contact library@up.edu.

Examining the Ways Female Students Experience
Teacher Feedback and Assessment

by

Nicole Foran

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Education
in
Learning and Leading

University of Portland
School of Education

2022

**Examining the Ways Female Students Experience
Teacher Feedback and Assessment**

by

Nicole Foran

This dissertation is completed as a partial requirement for the Doctor of Education (EdD) degree at the University of Portland in Portland, Oregon.

Approved: _____
REDACTED

Chairperson
REDACTED

Date 3/28/2022

Committee Member [initials]
REDACTED

Date 3/28/2022

Committee Member

Date 3/28/2022

If applicable:

Additional Committee Member

Date

Additional Committee Member

Date

Approved: REDACTED

Graduate Program Director
REDACTED

Date 3-28-22

Dean of the Unit
REDACTED

Date 3/28/22

Dean of the Graduate School or Representative

Date 3-28-22

Abstract

Providing instructional feedback and assessing student learning are integral components of the educational process. Yet teachers report that in their preservice education programs and subsequent inservice professional development, they receive very little instruction specifically dedicated to feedback and assessment practices. Research reveals that feedback and assessment remain at the center of debate in academia, and teachers report that assigning grades and giving effective academic feedback are major contributors to overall job dissatisfaction, especially given the high-stakes role grades play in students' academic, athletic, college admission, and future employment opportunities (Feldman, 2019; Guskey & Bailey, 2001). While much has been learned about effective assessment practices, there is still much to learn about how students experience feedback and grades. The pressure to achieve impacts students' motivation to learn, academic self-efficacy, self-esteem, and overall health and wellness, and female adolescents are at greater risk of mental-health issues due to grades and academic pressure than their male counterparts (Saviola et al., 2020; Villeneuve et al., 2019). The purpose of this mixed-methods case study was to better understand the effect of feedback and assessment practices on female students in an urban, faith-based environment. In phase 1 of this research, two surveys were distributed electronically. The first survey asked teachers to rate their self-perceived skills with assessment practices. The second survey asked students to describe how

they respond to teachers' instructional feedback. The second phase of this research involved semi-structured interviews where ten students described their experience with teacher feedback specific to an assignment they chose. The findings of this study were summarized in three areas: (1) teacher training in assessment versus their perception of assessment skill, (2) giving academic feedback versus experiencing academic feedback, and (3) considering students' emotional connection to their work in overall feedback and assessment strategies. This study contributed to our understanding of teachers' formal training versus perceived skills in feedback and assessment, how female students experience feedback, and the impact that teachers' instructional comments and final grades have on female students' academic self-efficacy and self-esteem. Keywords: feedback, assessment, grades, equitable grading practices, all-girls high school, female students, Catholic school, self-efficacy, self-esteem, perfectionism, academic identity, secondary assessment, teachers' comments, qualitative mixed methods

Acknowledgments

I am deeply grateful to my dissertation committee chair, Dr. Randy Hetherington, for his patient and insightful mentorship, and to my committee members, Dr. Nicole Ralston and Dr. Jaqueline Waggoner, for their guidance and unwavering support. All students should be so fortunate to experience the attention to relationship, the gift of encouragement, and the generosity of time and spirit that were afforded to me by these three incredible professors and that are the very subject of this work.

I am deeply grateful to my children whose love and encouragement lift me up and make me believe that I've done something incredibly worthwhile in my life. And I am deeply, profoundly, madly grateful to my husband, my intellectual partner, who understands the power of positive feedback, and who never tells me who I should be, what I should be, or how I should get there—but instead loves me for everything I am (and who would hate that I used an “m” dash in this sentence and that makes me love him even more). He is my greatest support and the finest teacher I have ever known.

Dedication

To the many students who inspired this work and who helped teach me that how they experience feedback matters.

Table of Contents

Abstract.....	iii
Acknowledgments	v
List of Tables	xi
Chapter 1: Introduction.....	1
COVID-19 Impacts on Students’ Learning and Social-Emotional Well-Being.....	2
Academic Repercussions of COVID-19 Related to Grades and Assessment.....	3
Opportunities Resulting from the Pandemic	4
Problematic Feedback and Assessment Practices.....	5
Impact on Girls More Significant	7
Purpose of the Study	8
Significance	9
Definition of Key Terms.....	10
Summary	12
Chapter 2: Literature Review	14
Function of Grades.....	15
History of Grading in the United States.....	16
Challenges with Assessment.....	17
Early Concerns	18
Contemporary Challenges	24
Impact of Assessment	34
Impact of Assessment Practices on Students’ Motivation to Learn.....	35
Impact of Assessment Practices on Students’ Social-Emotional Health	37
Impact of Assessment Practices in Catholic High Schools.....	39
Teacher Feedback.....	40
Females’ Experiences with Assessment and Feedback	42
Impact on Female Self-Esteem	44
Impact on Female Academic Motivation and Self-Efficacy.....	44
Summary.....	47

Chapter 3: Methodology.....	49
Research Questions.....	49
Rationale for Methodology.....	49
Context.....	50
Description of Setting	51
Participants	52
Design and Procedures.....	54
Phase 1	54
Phase 2	55
Instruments.....	57
Phase 1: Faculty Survey Instrument.....	57
Phase 1: Student Survey Instrument	60
Phase 2: Student Semi-Structured Interviews	62
Role of the Researcher	64
Data Analysis	65
Quantitative Analysis	65
Qualitative Analysis	66
Connecting the Data	67
Ethical Considerations	67
Ensuring Quality	67
Truth Value	68
Applicability and Consistency	68
Neutrality.....	68
Summary	69
Chapter 4: Data & Analysis.....	70
Research Question #1: Teachers’ Self-Perceived Skills with Feedback and Assessment.....	70
Scoring the STAP Survey	71
Relationship between Teachers’ Experience and STAP Score.....	72
Selecting and Developing Assessment Methods.....	73
Administering, Scoring, and Interpreting Assessment Results.....	75
Using Assessment Results to Inform Day-to-Day Decisions	77

Communicating Assessment Results to Others.....	79
Ethical Use of Assessment	81
Overall Skill with Assessment Practices.....	83
Summary of Research Question #1	85
Research Question #2: Female Students’ Experience with Academic Feedback.....	87
IFOS: Utility of Grades and Academic Feedback.....	87
IFOS: Sensitivity to Grades and Academic Feedback	89
IFOS: Importance of Confidentiality with Grades and Academic Feedback	91
IFOS: Retention of Grades and Academic Feedback	93
Student Interviews.....	95
Summary of Research Question #2.....	103
Summary	103
Chapter 5: Discussion.....	105
Discussion of Findings.....	106
Teacher Training in Assessment vs. Their Perception of Assessment Skill	106
Giving Academic Feedback vs. Experiencing Academic Feedback: Intent vs. Impact.....	108
Considering Students’ Emotional Connection to Their Work in Overall Feedback and Assessment Strategies	109
Conclusion of Findings	110
Limitations of the Study	112
Limitations Due to COVID-19 Pandemic.....	112
Limitations Due to Role of Researcher	113
Limitations Due to “Neutral” Option on Student Survey	113
Limitations Due to Small Sample Size of Student Interviews	116
Limitations Due to Reliability and Validity of Teachers’ Self-Reporting.....	116
Recommendations.....	117
Implications for Future Research.....	118
Conclusions.....	119
References	122
Appendix A	133
Appendix B.....	137

Appendix C.....	140
Appendix D	142
Appendix E.....	143

List of Tables

Table 1. Timeline of Procedures	56
Table 2. Instrument Alignment with Phases of the Study	57
Table 3. Research Question #1 - Faculty Survey: STAP Assessment Areas by Domain and Question	59
Table 4. Demographic Characteristics of Participants	71
Table 5. Relationship between Years Teaching/Total STAP Score	73
Table 6. Teachers' Self-Perceptions of Their Knowledge, Experience, and Skills with Selecting and Developing Assessment Methods	74
Table 7. Teachers' Self-Perceptions of Their Knowledge, Experience, and Skills with Administering, Scoring, and Interpreting Assessment Results	76
Table 8. Teachers' Self-Perceptions of Their Knowledge, Experience, and Skills Using Assessment Results to Inform Day-to-Day Decisions	78
Table 9. Teachers' Self-Perceptions of Their Knowledge, Experience, and Skills with Communicating Assessment Results to Others	80
Table 10. Teachers' Self-Perceptions of Their Knowledge, Experience, and Skills with Ethical Use of Assessment	82
Table 11. Teachers' Self-Perceptions of Their Overall Skill with Assessment Practices	84
Table 12. STAP Items by Domain	86
Table 13. Students' Perception of Feedback Utility	88
Table 14. Importance of Considering Students' Sensitivity to Feedback	90
Table 15. Importance of Confidentiality in Receiving Academic Feedback	92
Table 16. Students' Retention of Academic Feedback	94
Table 17. Themes and Categories in Second-Cycle Coding Analysis	97

Table 18. Responses to Question: “Tell Me Why You Chose This Assignment to Talk About.....	98
Table 19. Responses to Questions about Academic Feedback.....	101
Table 20. Treatment of “Neutral” Data When Considering Students’ Sensitivity to Feedback.....	115

Chapter 1: Introduction

Feedback and assessment, as practices designed to improve student achievement and inform teaching decisions, remain at the center of debate in academia (Feldman, 2019; Guskey, 2011). Assigning grades and giving academic feedback involve some degree of subjectivity, and even with training, teachers often resort to grading practices they experienced as students and that reflect their own views of assessment (Guskey & Bailey, 2001). Even in a best-case scenario where teachers might receive preservice and ongoing inservice specifically focused on feedback and assessment, these practices, by their very nature, are impacted by the context and medium of delivery and disparities and discrepancies in outcome. While much has been learned about the effectiveness of assessment tools and strategies and the norming of evaluative practices, there is still much to learn about how students experience feedback and assessment.

Educators and students find grading disparities and discrepancies frustrating, and the stakes are high for today's students as grades play a critical role in their academic and athletic opportunities, as well as their future college admission and employment (Feldman, 2019). And students feel this pressure. A recent study reveals that students are experiencing greater anxiety and stress, a mental state experienced by adolescents and characterized by an overwhelming sense of worry, apprehension, and tension related to school, and studies are pointing to grades as a key contributor (Saviola et al., 2020; Villeneuve et al., 2019). In fact, a high percentage of students who attend high-performance—or college-preparatory—high schools report stress, exhaustion, and loss of sleep due to prioritizing academic outcomes over their own

health and well-being; research also reveals that girls are at a greater risk than boys to experience these mental health and physical responses (Altermatt, 2015; Twenge et al., 2019; Villeneuve et al., 2019). Thus, the type of feedback, how it is given, and the way students experience it matters.

COVID-19 Impacts on Students' Learning and Social-Emotional Well-Being

The events of Spring 2020 disrupted many elements of traditional American life, and perhaps most markedly, the ways in which schools delivered education to schoolchildren of all ages. With the onset of the novel coronavirus (COVID-19) pandemic, students—abruptly separated from their peers and the familiar context of their daily school lives—engaged with teachers virtually in spaces such as bedrooms, living rooms, and kitchens (Feldman, 2020). Students and teachers struggled to find their footing, and pediatricians reported higher rates of adolescent depression, anxiety, and suicidal ideation resulting from the dramatic increase in social isolation (Marques de Miranda et al., 2020). In 2020, researchers Magson et al. surveyed 248 adolescents in Canada, China, Australia, and the United States to determine the psychological impact of the COVID-19 pandemic. They gathered data before the transition to online learning and then again two months following the government-implemented lockdowns in each country. Early indications from this study reveal a consistent increase in anxiety and depression, as well as an overall decrease in life satisfaction. While it is too early to determine the extent the pandemic affected adolescents' overall well-being, this early study reveals a consistent link between the pandemic and the marked decline of the mental health of adolescents—circumstances that make assessing students' learning even more difficult (Magson et al., 2020).

Academic Repercussions of COVID-19 Related to Grades and Assessment

Along with the concerns over social and emotional health and well-being revealed by the Magson et al. (2020) study, researchers are just beginning to understand the potential long-term academic repercussions of COVID-19 on today's students (Magson et al., 2020; Marques de Miranda et al., 2020; Sawchuck, 2020). To contain the spread of the coronavirus in the United States, ever-changing restrictions impacting schools had school leaders grappling with how to abruptly transition the overall educational experience—experience that typically includes instruction, feedback and assessment, academic and personal support, access to nutrition, and interpersonal connection—to online learning. Pandemic-mandated transitions in K-12 education highlighted disparities between students' access to reliable technology, students' ability to navigate the online environment, and students' ability to learn in relative isolation from their peers (Feldman, 2020). Students' home and socio-economic circumstances vary greatly, and there were considerable discrepancies in available personal and academic supports among individual schools and school districts. Some schools transitioned to online learning offering economic and academic supports for struggling students, as well as opportunities for important virtual interactions and social connection. However, even these schools struggled with how best to assess student learning. Some schools awarded course credit if a student simply showed up for the five-minute virtual attendance portion of a class period, even awarding an A on an academic transcript to students who turned in little to no work (Anand & Bhatia, 2021; Sawchuck, 2020). Other schools assigned letter grades using prepandemic grading policies, and some were willing to fail students not meeting the

minimum course requirements (Sawchuck, 2020). Assessment proved to be a major source of stress for students and teachers alike, and the pandemic helped magnify what many argue to be the flawed and inequitable practices that have plagued grading and assessment for decades.

Opportunities Resulting from the Pandemic

While it may take years to understand the negative impacts of increased isolation and online learning resulting from the COVID-19 pandemic, the current context has provided the impetus to reevaluate traditional educational practices based on the disparities and discrepancies indicated above. Arguably, the fundamental objective of educators is to help students succeed, yet for many students, parents, and teachers in the wake of the pandemic, the goals may have shifted to prioritizing emotional health and well-being (Feldman, 2020; Marques de Miranda et al., 2020). Studies show that stress interferes with deep learning and the brain's ability to process, recall, and perform high-level cognitive tasks, which further complicates the re-entry of students to traditional, prepandemic school settings (Feldman, 2020). Increased levels of stress and anxiety in both students and teachers can result from transitions, adapting to new technologies, lack of in-person schooling, and social isolation (Sawchuck, 2020; Scott et al., 2021). These are all aspects of learning in the COVID-19 era. Perhaps now more than ever, educators and educational leaders must examine current educational practices that contribute to student stress and anxiety. The 2021-22 school year will provide the backdrop for reevaluating common feedback and assessment practices used by classroom teachers—practices that many researchers

argue are inconsistent, inequitable, and even harmful to students' health and well-being (Feldman, 2019, 2020; Marques de Miranda et al., 2020).

Problematic Feedback and Assessment Practices

One of the fundamental elements of a teacher's role is to help students achieve, and research underscores that the measuring of students' learning and progress is an essential part of the achievement process (Feldman, 2019; Guskey, 2011; Guskey & Bailey, 2001). Few aspects of education are as controversial as assessment, a practice that involves measuring student learning and reporting the results. Research reveals that traditional feedback and assessment practices are often inconsistent, biased, and inequitable, and the anxiety students experience associated with assessment was on the increase even before the onset of the COVID-19 pandemic in March 2020 and the subsequent transition to online learning in the United States (Feldman, 2020; Villeneuve et al., 2019). Students are not the only ones experiencing stress and anxiety around classroom assessment practices. The assigning of grades and giving effective academic feedback remain areas where teachers express job dissatisfaction, frustration, and confusion, and some teachers point to the lack of preservice and inservice training as major contributors to the problem (Feldman, 2019). Because research suggests that teachers experience stress and anxiety with identifying and implementing assessment practices that inform teaching and learning and that support student well-being, and students experience increased stress and anxiety related to traditional grading practices, it is critical to reevaluate assessment practices, to understand the impact on students' motivation to learn, academic self-efficacy, and social and emotional well-being. These factors all contribute to students' ability to

achieve in the K-12 system (Guskey, 2011; Pomerantz et al., 2002; Villeneuve et al., 2019).

Reevaluation should consider research on grading practices that underscores the inherent inconsistencies and inequities with assessment as it is currently practiced in many schools. The subjective nature of assigning grades, the overall lack of training for teachers with regards to assessment, the inclusion of student behaviors not related to measuring student learning, and the implicit biases connected to the practice of grading frustrate and confuse both teachers and students (Feldman, 2019; Guskey & Bailey, 2001; Marzano, 2000) and contribute to students' stress and anxiety levels (Marques de Miranda et al., 2020; Sawchuck, 2020). These inconsistencies and inequities related to grading often do more than produce unreliable results. Brookhart (2013) suggests the detrimental impact of common grading practices can impair students' intrinsic motivation to learn and achieve, arguably the very reason for schooling in the first place. Moreover, the results of common assessment practices—final grades—can be used by students, parents, colleges and universities, and employers to make decisions about educational opportunities. Additional studies reveal that formative and summative grades impact students' opportunities while they are in school and beyond, potentially influencing career choice, career satisfaction, and earning potential (Feldman, 2020; Johnson & Helgeson, 2002; Rinfret et al., 2014). Grades create lifelong achievement and opportunity gaps, gaps perpetuated from generation to generation, especially for female students and students of color (Altermatt & Pomerantz, 2003; Feldman, 2019, 2020; Pomerantz et al., 2002).

Impact on Girls More Significant

Research suggests that girls are influenced to a higher degree by others' feedback than boys, and girls suffer from anxiety and depression at higher rates than boys (Altermatt et al., 2003; Costello et al., 2006; Mechtenberg, 2009; Mohanty et al., 2015). The COVID-19 pandemic brought about a sudden transition to online learning, a sporadic return to in-person learning, and a continual threat of returning to online learning with each new coronavirus variant. The stress and anxiety these transitions create for all students amplify the critical need to reevaluate current grading practices, especially for girls who are already predisposed to depression and anxiety (Feldman, 2020; Pomerantz et al., 2002). With the elements of the educational experience in flux as students and teachers transition back to a more traditional learning experience, the time is right for teachers and school leaders to examine current feedback and assessment practices, and prioritize changes that promote student learning and wellness.

According to a Pew Research Center study (2019) that examined the pressures on today's adolescents, anxiety and depression emerged as top concerns, with the pressure to do well in school topping the list of major contributors to their stress. For all students, and especially those students who struggle, the pressure associated with academic achievement can create levels of anxiety that not only impede their learning, but also threaten their emotional health and wellbeing (Feldman, 2020; Pomerantz et al., 2002). Creating equitable assessment practices that address the needs of all students is difficult, and research suggests that it is complicated further by the differences in how females and males experience feedback and assessment practices

(Altermatt & Pomerantz, 2003; Johnson & Helgeson, 2002). In fact, research suggests that compared to males, adolescent females experience a significantly higher level of academic anxiety, the sense of worry, apprehension, and tension related to school (Altermatt & Pomerantz, 2003; Saviola et al., 2020). One reason may be that because boys are less concerned with pleasing adults, they experience less internal distress caused by the pressure to achieve in school (Pomerantz et al., 2002). Because girls are more motivated to please adults, this added internal distress and academic anxiety may influence their overall achievement, their intrinsic desire to learn, and their academic self-efficacy (Bandura, 1997; Pomerantz et al., 2002; Schunk & Pajares, 2005). Thus, it is critical that educators understand how girls experience elements of the school experience, including assessment (Johnson & Helgeson, 2002; Mohanty and Jena, 2015). While evaluative feedback and grades can be essential components in motivating students to higher levels of learning and a greater sense of academic self-efficacy (Guskey, 2011; Guskey & Bailey, 2001), feedback and assessment practices that fail to prioritize student learning may negatively affect girls' self-efficacy to a greater degree. Currently, there is a paucity of research on how female high school students experience feedback and assessment practices, and specifically how teacher feedback impacts their self-efficacy to achieve. This study seeks to fill that gap.

Purpose of the Study

The purpose of this case study was to understand the impact of feedback and assessment practices on female students in an all-girls high school in an urban, faith-based environment. The specific research questions guiding this study were:

1. What do teachers at an all-girls, faith-based urban high school perceive to be formative and summative assessment strategies that support improved student learning in their classroom?
2. How do female students experience feedback and assessment practices in an all-girls, faith-based, urban high school?

Significance

Assessment practices in high schools and how the results of those assessments are communicated to students, parents, colleges and universities, and employers can result in inequitable outcomes for students. In addition to the inequities inherent in assessment practices in public or charter schools, assessment in Catholic high schools presents a unique set of struggles concerning equity (Imperial, 2012). Before 1973, Catholic high schools in the United States, schools founded on a mission of serving poor and marginalized members of society, had done little to welcome or support students with diverse needs in their communities (McDermott, 1997). In 1973, this changed when the United States Conference of Catholic Bishops published its pastoral message, *To Teach as Jesus Did* (U.S.C.C.B., 2002). This meeting of bishops and the resulting pastoral statement required Catholic schools to confront their exclusionary practices and commit to creating school environments that mirrored the mission of the Church. This action expanded Catholic-school communities to include students with learning differences, students who struggled academically, and students who came from socially and economically diverse backgrounds. It required schools to research current methodologies on how best to meet the needs of all students served by the school including assessment practices (Imperial, 2012). There appears to be a dearth

of literature directly addressing grading practices in Catholic high schools, but in 2012, researcher Peter Imperial studied assessment in Catholic schools to bring the voice of Catholic educators into the conversation. Using random sampling, he surveyed 486 Catholic secondary teachers and 50 administrators from 33 high schools in California, Nevada, and Hawaii. He found that, much like the research on grading in traditional high schools (Brookhart, 2008; Guskey, 2011), Catholic high-school teachers included a combination of achievement evidence and non-academic evidence to determine students' grades (Imperial, 2012). For example, Imperial (2012) found that over 78% of the teachers in the study used formative assessments to determine students' final grades. Formative assessments used in this manner invite bias, result in inaccurate descriptions of student learning, and motivate students to be more concerned about their grades than their learning (Feldman, 2019, 2020). These assessment practices combined with the often-unrelenting pressure to succeed that can emanate from society, family, or self (Madigan, 2019; Villeneuve et al., 2019) can contribute to increased levels of anxiety in young people, especially girls (Altermatt & Pomerantz, 2003). This should be concerning to pediatricians, parents, and educators alike.

Definition of Key Terms

There are many terms used regarding assessment and grading, and they are often defined in many ways depending on the person(s) and context in which they are being used. The following definitions about assessment, grading, school organization, and perceived academic ability are provided for clarity and apply to this study.

Assessment. The process of gathering data about student achievement. Teachers, schools, school districts, and state and federal governments use this information to make decisions (Marzano, 2010; Nitko & Brookhart, 2011).

Assessment: Formative. A judgment on the progress and quality of student achievement determined while the student is in the process of learning. This type of assessment informs both the student and teacher (Nitko & Brookhart, 2011).

Assessment: Summative. The determination of student achievement at the end of the instructional process, such as the end of a unit or the end of a grading term (Nitko & Brookhart, 2011).

Feedback. The information given to students that describes their academic performance as well as non-achievement related behaviors. Feedback may or may not include a grade and is intended to promote academic improvement (Poulos & Mahony, 2008).

Grades. A standardized system of letters, numbers, or numerical percentages that communicate student achievement at the end of a set period. Grades indicate the results of summative assessments (Feldman, 2019; Nitko & Brookhart, 2011).

Self-Efficacy. One's perceived capabilities to learn or perform behaviors at a designated level and to a specific degree (Bandura, 1997; Schunk & Pajares, 2005).

Single-Sex Schools. Elementary, secondary, or post-secondary schools in which females or males attend exclusively with members who identify with their sex cohort (Mael et al., 2005).

Standards-Based Grading System. An approach to grading where course goals and criteria are clearly defined and students are evaluated solely on their

proficiency against these criteria. In standards-based grading, only proficiency matters, and students' attempts at proficiency are not included in the final grade. In addition, non-academic elements, such as participation or work habits, typically are excluded or have minimal impact on final grade (Marzano, 2010).

Traditional Grading System. The combination of formative and summative assessments to determine a numerical percentage or letter grade (A-F) (Nitko & Brookhart, 2011). Traditional grading practices may include tracking points for homework and assessments, use weighted categories and achievement curves, and include assessment of non-achievement behaviors in final grades (Feldman, 2019).

Summary

The research is clear that feedback and assessment practices impact student learning, motivation to learn, academic self-efficacy, and social and emotional well-being (Brookhart, 2013; Feldman, 2019; Guskey, 2007; Rosenberg & Simmons, 1975; Villeneuve et al., 2019). Given that males and females experience these impacts differently, and that there is little research related to feedback and assessment in Catholic high schools, it is important to understand how girls in an all-girls, faith-based, urban high school experience feedback and assessment. A deep understanding of teacher assessment and feedback practices may highlight the potential disparity between the intent of feedback and how the feedback is interpreted and experienced. For the teachers, knowing the realities of how students experience assessment informs opportunities for refining practice and improving outcomes. This study will provide data that can inform teacher leaders' and administrators' decisions on inservice opportunities focused on feedback and assessment practices that are equitable and

meet the needs of female students attending Catholic high schools. Schools designed to empower young women should employ best practices identified by students as supportive of their goals and that do not contribute to increased levels of stress or anxiety.

Chapter 2 of this study provides an overview of the literature on feedback and assessment practices, the challenges and limitations of these practices, the essential role of feedback in academic self-concept and academic achievement, and how girls, specifically, experience these practices. Chapter 3 will outline the methodology used in this qualitative mixed-methods case study design (Yin, 2018) and the method of data collection and analysis. Chapter 4 will describe the study's findings, and Chapter 5 will outline the analysis of the findings, provide the limitations of the work, and conclude with recommendations for the profession including topics requiring further research arising from the work.

Chapter 2: Literature Review

Reviewing the literature on academic feedback and assessment practices in 21st-century schools to gain a deeper understanding of underlying principles is a necessary precursor to learning how students experience these practices, specifically female students in an urban, all-girls, faith-based school. For many educational leaders, an important step towards making informed decisions about which current practices are best practices is to first take inventory of current practices and explore the extent to which teachers feel confident using the various feedback and assessment tools to instruct students on how to improve and to measure learning (King et al., 2009; Sheen, 2004). Results of the literature review and student interviews within this study may help inform the school's leadership about perceived problems with current feedback and grading practices from the student perspective; illuminate the extent these practices may influence students' motivation to learn, academic self-efficacy, and overall health and well-being; and, as the result of further reflection and discernment by teachers and leaders, precipitate a review of the school's assessment practices and provide focus areas for future professional development and school improvement planning.

This section provides an overview of assessment practices and how female students experience academic feedback and grades. The review begins by identifying the primary function of grades in the learning process as documented in the literature and provides a brief history of grading practices in the United States. Next, concerns with grading practices and the challenges inherent to grading and reporting will be addressed, as well as research related to the impact of feedback and grading practices

on students' motivation to learn and social-emotional health and well-being. Research identifying specific differences in the ways girls experience common feedback and grading practices will conclude the review.

Function of Grades

There is considerable discussion in the literature about the overall function of grading and reporting and the role assessment plays in student learning (Guskey, 1996; Black & Wiliam, 1998). In Thomas Guskey's (1996) seminal article "Reporting on Student Learning: Lessons from the Past—Prescriptions for the Future," he chronicles a history of grading in the United States and highlights the inherent tension with teachers having to play the roles of both student advocate and judge. British professors Black and Wiliam (1998) continue the discussion in their influential piece *Assessment and Classroom Learning* and differentiate between formative and summative assessment. They argue that while classroom assessments often are to blame for creating a culture of competition and superficial learning, assessments designed to give feedback during the learning process foster critical thinking and achievement (Black & Wiliam, 1998). A common refrain in both articles is that academic feedback and assessment remain part of the learning experience, and the evaluation of student learning—while challenging and often problematic—continues to inform important instructional and administrative decisions. Although assessing student learning has been a part of traditional schooling since the inception of the first schoolhouse, the concept of formal grading and reporting are comparatively recent phenomena (Guskey, 1996). Most current grading practices were developed more than a century ago, yet these practices according to recent scholars have different implications for

today's young people (Feldman, 2019, 2020; Villeneuve et al., 2019). To understand the debate that often surrounds contemporary grading and reporting, it is necessary to trace the history from the first assessments, narrative feedback reports, to common 21st century assessment practices.

History of Grading in the United States

The origins of grading in the United States can be traced to the ancient Greeks who conducted oral exams where students demonstrated the extent to which they mastered the content and their eloquence in expressing their learning (Guskey, 1996). These narrative exams were informal evaluations that provided teachers with data about areas where students needed additional instruction. European universities dating back to the 12th century followed suit and modeled their assessment of student learning after these early narrative reports (Brookhart, 2008). Grading in American schools can be traced to American colleges in the 18th century that likely imitated the systems and philosophies of their European predecessors (Durm, 1993). Before 1850, grading and reporting practices in common schools were less important as few students progressed beyond the primary grades, or the first four or five years of education, which took place in one-room schoolhouses (Guskey, 1996). As the number of students going beyond the primary grades increased toward the latter part of the century, schools began to organize students according to age. At that point, teachers created written records of student progress, and students were required to demonstrate proficiency of skills in order to progress to the next grade level (Guskey, 1996).

The number of students enrolled in elementary schools increased substantially between 1870 and 1915, and the number of students entering high schools increased from 500 to 10,000 (Gutek, 1986). By 1900, 32 states had passed compulsory education laws. With more students and an increasingly diverse student population, high schools shifted from assessing proficiency in skills using written narrative descriptions to assessing students' accomplishments in discrete subject areas employing percentages and, eventually, alphanumeric grades that included letters representing a range of achievement (Brookhart, 2008). Yale University implemented a system that categorized student performance using a four-point scale; Harvard University adapted this into six distinct percentage "divisions"; and, in 1897, Mount Holyoke College implemented letters grades (A-E) to communicate student performance (Guskey, 1996; Marzano, 2000; Vatterott, 2015). In the early 1900s, most American high schools used percentage grading to communicate student performance (Guskey, 1996; Vatterott, 2015). With individual schools now interpreting and adapting grading models and teachers interpreting and employing these various systems, grades became less reliable indicators of student achievement (Guskey, 1996; Starch & Elliott, 1912).

Challenges with Assessment

There are many challenges facing teachers and students in terms of assessment practices. These challenges stem from early concerns related to the shift from local, one-room and one-teacher schoolhouses to schools with professional educators organized into regional districts. This changing landscape allowed for inconsistent grading practices to take hold, as teachers had varying levels of training and different

perspectives on how to assess learning. Teachers and students in today's schools continue to face challenges with assessment, and these challenges will be explored as to what the literature has already uncovered.

Early Concerns

Comprehensive evaluations that primarily communicated student progress and areas for growth and improvement began to shift to a more cursory, expedient system of assessment (Guskey, 1996). This shift marked the beginning of contemporary grading systems in American schools, a shift that coincided with changes in the structure of schools as well as the population of students that schools now served (Guskey, 1996; Guskey & Bailey, 2001). Until about the turn of the 20th century, schools were locally controlled, had one teacher (usually a female who stayed until she married), and served students from different age groups (Brookhart, 2013). Students demonstrated their learning with high-stakes, end-of-year recitations, a practice that placed value on a student's ability to memorize material and perform on demand (Brookhart, 2013; Haertel & Herman, 2005). The structure of schooling began to shift from locally controlled schools to regional school districts where a district included lower schools (various configurations of today's grade schools and middle schools) and a high school (Brookhart, 2013). In addition, students were now taught by professional educators, as opposed to before when teachers were predominantly inexperienced (Brookhart, 2013). Because the shift was gradual, few American educators sounded the alarm when inconsistent practices among professional educators with different levels of training, experience, and expectations began to emerge.

Inconsistent Practices. The literature reveals that many assessment strategies used throughout much of the 20th century in America differed in their purpose or intent as well as the criteria being applied. In 1912, two professors from the University of Wisconsin, Daniel Starch and Edward Elliott, looked at teacher subjectivity in grading students' English written work. This seminal study exposed the extent to which subjectivity—personal values and expectations—influenced teachers' grading practices (Starch & Elliott, 1912). In this study, they distributed copies of exam papers written by two different freshmen students to freshmen English teachers at 200 high schools. Teachers were asked to grade the papers according to their school standards on a 100-point scale (where 75 was considered passing) and 142 schools of the 200 participated. The scores on one paper ranged from 64 to 98, and the other paper's scores varied from 50 to 97. Starch and Elliott (1912) also found a wide array of inconsistencies. While some teachers were influenced by factors like punctuation, spelling, and overall neatness in the final grade, others gave more importance to elements such as grammar and style (Starch & Elliott, 1912). They concluded that because grading is subjective, grades for any paper were unfair and unreliable. This study created controversy within the academic community (Brookhart, 2013). Some argued that the study was inherently flawed because assessing student writing is, by its nature, a subjective practice, and a variation in scores should be expected (Guskey, 1996). To counter this argument, Starch and Elliott conducted a similar study in 1913, this time using geometry exams, and the results revealed an even greater variation in score (Starch & Elliott, 1913). In fact, some teachers deducted points for incorrect answers, while others, like the teachers scoring the English exams, allowed neatness,

form, and spelling to influence students' final scores. To ensure a fairer, more equitable distribution of grades and to remedy the unreliability that results from subjectivity, Starch advocated for a standardized grading scale with fewer categories (Starch, 1916). But the debate surrounding the reliability of assessment, even with more accountability and standardization, was far from over.

Challenges Defining the Purpose of Assessment. While the discussion on the reliability and overall value of assessment continues today, there appears to be consensus among educational leaders on some key elements, such as there is not one method of assessment that, by itself, provides a comprehensive look at student achievement; and assessment will usually involve some degree of subjectivity (Guskey, 1996; Marzano, 2010). Yet the actual purpose of grading and reporting lacks consensus.

Some view assessment not as a valuable source of information used to improve student learning, but as a tool for summarizing learning for teachers to use in grade reports and even for labeling and ranking students for school reporting purposes (Heritage, 2007). To respond to these types of claims, researcher Thomas Guskey (1996) distilled the purpose of grading into six categories:

- 1) Communicating student achievement
- 2) Providing incentive for student learning
- 3) Providing students information for self-evaluation
- 4) Helping students determine educational paths and programs
- 5) Evaluating the effectiveness of instructional methods and programs
- 6) Providing evidence for lack of achievement (p. 17)

In 2007, Thomas Guskey expanded on the previous article (Guskey, 1996) to argue that the goals of assessment fall short, especially since summative grades include non-academic factors. Further, Guskey (2007) contends that neither administrators nor teachers perceive these end-of-course grades to be a particularly accurate gauge of student learning. While grades can provide important feedback about student achievement, studies reveal that many teachers fail to employ practices that elicit accurate results (Feldman, 2019; Guskey & Bailey, 2001; Marzano, 2000). In fact, literature suggests that little has changed since Starch and Elliott's studies on grading in both English and math revealed the subjective element to grades and the inability for one grade to capture the nuances of achievement (Starch & Elliott, 1912, 1913). Yet despite Starch and Elliott's seminal studies on assessment, grading and reporting—the accepted mechanism for communicating student progress across the United States and arguably many nations across the globe—has remained virtually unchanged for over a century (Starch & Elliott, 1913; Starch, 1916; Vatterott, 2015).

Challenges Due to Changing Demands of Schools. Grading practices were becoming even less effective with the changing demands of modern schools—demands that now required schools to serve greater numbers of students including students from a wider range of cultural and socioeconomic backgrounds and abilities (Feldman, 2019). Joseph Feldman, who writes extensively about inconsistent and unfair grading practices, describes the inherent flaws of a grading system designed as an efficient sorting system (Feldman, 2019). The increasing numbers of students combined with decreasing financial resources to meet their diverse needs required more efficient, homogenous grading systems that were manageable given increased

class sizes (Feldman, 2019). Many 20th century schools began implementing the A-F grading scale that allowed for more expedient and often less-specific descriptions of student achievement than points and percentages (Guskey, 2011). The pressure for developing a more standardized system served a bureaucratic function as well. Standardized systems were a recognizable way to signify student achievement for external audiences, including colleges and employers (Craig, 2011).

Together with adopting an A-F grading system to evaluate and sort students, 20th century schools adhered to the common notion of the time that there is a bell curve, or normal distribution of intelligence, within a population, a theory introduced by Alfred Binet early in the 20th century (Feldman, 2019). Although widely debated today, Binet's concept of natural intelligence—one's innate and immutable mental ability—became popular throughout the last century and paved the way for assessments meant to sort and categorize, such as the Intelligence Quotient (I.Q.) test (Feldman, 2019; Guskey, 2011). This concept of sorting and distribution was replicated in academia giving rise to the grading curve, a method of grading that forces teachers to assign a predetermined number of grades to resemble the distribution of a bell-shaped curve (Guskey, 2011). Many schools today still employ the grading curve, but much like I.Q. tests and curved grading, many standardized tests have been shown to be biased and serve to reinforce social inequities associated with oppression (Feldman, 2019, 2020). Some argue that while the I.Q. test was developed to measure individuals' intelligence, the original intent really was to secure opportunities for highly intelligent people (Lemann, 1997). Feldman (2019) argues that norm-referenced grading—or “grading on a curve”—simulates this same unjust

phenomenon. Grading on a curve provided opportunities for inaccuracy as teachers no longer needed to assess student progress against specific learning criteria, and instead assigned a grade based upon the student's achievement relative to the achievement of other students (Feldman, 2019). This practice created two differing schools of thought: one that believed the accuracy and objectivity of norm-referenced testing could expand educational opportunities and one that believed this type of testing limited student opportunities by reinforcing inequitable societal hierarchies (Feldman, 2019).

Lack of Agreement in Assessment Philosophy. By the early 1960s, assessment practices meant to evaluate student progress in relation to other students' progress were beginning to lose ground in favor of a new philosophy that now focused on objective criteria to assess student work and measure individual achievement (Brookhart, 2008). This shift in philosophy paved the way for the Standards Movement of the 1980s, a national education strategy that attempted to improve consistency of the scope and sequence in curriculum and instruction and reduce teacher subjectivity in grading and reporting (Cox, 2011). The Standards Movement specified what content students should know in the core subjects at each grade level, described assessments that measured students' progress related to the standards, and implemented accountability systems that enabled schools, districts, and states to report progress against these goals (Schwartz et al., 2000). Iterations of this movement, such as No Child Left Behind in 2001 and the more recent Common Core State Standards, shifted the focus from assessment that relies on points, percentages, and the A-F scale to assessment based on identified objective criteria for measuring proficiency (Marzano, 2010). This shift arguably put students in charge of their growth and

learning and left fewer opportunities for teacher subjectivity concerning curriculum choices and learning assessment (Battistone et al., 2019). However, to be able to carry out the goals of the Standards Movement, required teachers to be both knowledgeable and conversant in the standards, plus highly skilled in determining whether students had demonstrated proficiency using identified objective criteria (Feldman, 2019). While standards-based grading should, by definition, make it easier for teachers to assign grades and for students to monitor their progress toward proficiency, this method can fall victim to the same shortcomings of traditional grading due to lack of teacher training and inconsistencies between teachers and schools (Battistone et al., 2019). Further, schools that reportedly adhere to standards-based grading where criteria are used for judging proficiency often include non-standards-based grading practices—participation, timeliness in submitting work, formative assessments, and other non-achievement-based assessment—in final grades; therefore, the labeling of these blended grading practices as standards-based grading is inaccurate (Feldman, 2019). Standards-based grading requires clear training and on-going professional development to ensure fidelity to its assessment philosophy, a challenge too great for many schools and districts (Battistone et al., 2019).

Contemporary Challenges

Because determining a grade can be complex and can vary by teacher, research has consistently identified the process as fraught with challenges (Guskey & Bailey, 2001). Teachers are tasked with grading student progress and achievement and reporting those grades to all stakeholders, including students, parents, schools, colleges and universities, and employers—stakeholders who make determinations

based on the results of these assessments. And teachers consider grading to be one of the least appealing and most unclear, stressful, and demanding facets of their jobs (Battistone et al., 2019). While teachers experience procedural, analytical, and ethical difficulties with grading, these difficulties can be even more pronounced for students who need to make sense of the unique, often idiosyncratic policies and practices of six or more teachers each term in the high school setting. These variable grading practices invite bias, provide misleading information, and create confusion for students and families (Feldman, 2020).

Today, shifting student demographics, a new focus on equity, and a century of study and research on grading has called traditional assessment practices into question (Feldman, 2019; Marzano, 2010; Tierney, 2015). Current assessment practices, while still adhering to traditional forms and functions, must reflect and respond to the needs of today's students. Educators now have a deeper understanding of how to respond to the rich cultural and ethnic diversity in schools and how students learn and experience traditional school settings differently (Feldman, 2019, 2020). Research also underscores the effects of trauma and anxiety for students, especially society's most vulnerable—students of color, students with special needs, and students with learning differences (Marques de Miranda et al., 2020; Saviola, et al., 2020; Sawchuck, 2020). This emergence of a more student-centered, culturally responsive, and differentiated approach to teaching, learning, assessing, and reporting, while purported to address many of the inequities, can do little to quell the controversies over grading (Feldman, 2019). Much of the current criticism of assessment focuses on the inconsistencies among teachers, the same criticism that Starch and Elliott raised in their 1912 and

1913 studies, but the impact of unfair, inconsistent, and biased grading practices may have greater implications for students today (Brookhart, 2013; Feldman, 2019, 2020; Guskey, 2007). Many argue that while the letters on a report card look the same as in the past, these letters potentially affect students' future opportunities, wreak havoc with their academic self-efficacy, and even alter the trajectory of their lives (Altermatt & Pomerantz, 2003; Feldman, 2019; Tierney, 2015).

Inconsistencies and Inequities. There are many potential reasons that inconsistencies and inequities continue in today's classrooms. In 1998, the College Board conducted a survey that reported 85% of schools nationwide allowed secondary teachers significant autonomy over establishing grading practices and norms for students. Battistone et al. (2019) noted that a study in 1998 that focused on early-career teachers' comfort and knowledge with assessment revealed the average level of experience for a teacher in an American public school was 15 years but, by 2013, the most common level of experience was only one year. Perhaps more alarming is that some teachers report their only formal training in assessment occurred during their preservice programs before they ever entered the classroom (Battistone et al., 2019). Battistone et al. (2019) explored the experience of 11 early-career teachers from four higher education preservice programs to determine how prepared they felt with their district's grading system (in this case, a standards-based grading system). The findings of this study revealed the teachers' common concerns about their preservice programs' training with assessment practices overall and concerns about the trustworthiness of the grades that they assign their own students (Battistone et al., 2019).

While the Battistone et al. (2019) study followed early-career teachers, early-career teachers are not alone in employing inconsistent and inequitable assessment practices. To explore inconsistent grading practices, one study compared teachers' perceptions of their students' learning gains over a two-year period with students' reported learning gains as measured by a standardized test (Ehrenberg et al., 1995). This study gathered data from 5,193 teachers in the first phase and 15,908 teachers in the second phase. Students were administered a cognitive test in their teacher's subject area, teachers were asked a set of questions about each of their students who were surveyed, and both students and teachers were asked a set of demographic questions. The findings suggest that the teacher's gender, ethnicity, and race played a substantial role in their evaluations of their students' learning, and the researchers concluded that determining the impact of a teacher's identity on the success of a student should be a high priority in determining ethical grading practices (Ehrenberg et al., 1995). The findings from this study did not suggest that a teacher's identity influenced gains in student learning, but the relationship between a teacher's identity and the impact to students' final grades might cause harm in other ways (Feldman, 2019). With the combination of evidence that early-career teachers feel insufficiently trained in assessment, that teachers' subjectivity and inherent biases influence assessment, and that the significant autonomy teachers have in developing assessments and scoring and reporting grades, it is no surprise that assessment is one of the most frequently identified areas for improvement in modern schools (Marzano, 2010).

Lack of Training, Clear Policies, and Oversight. In the 1998 College Board examination of high-school grading policies, only 7% of schools had in place general

guidelines for grading policies and only 4% had strict, clearly outlined grading policies. In this report, the College Board described the difficulty in evaluating students' grades given the wide latitude and flexibility schools allowed teachers. With great autonomy, teachers are tasked with making ethical judgments related to assessment, but they often lack formal training, or their training is no longer current (Green et al., 2007). Grade variability, a result of teaching autonomy and intellectual freedom, is problematic for various reasons, one of which is that often it leads to inequitable outcomes (Feldman, 2019). Creating uniform grading practices seems highly unlikely given that few resources discuss the general ethics of classroom assessment practices, much less the specifics (Green et al., 2007). In a 2007 study that surveyed preservice teachers, inservice teachers, and administrators enrolled in an assessment course at two large southern universities, participants were asked to judge and label a variety of grading scenarios as ethical or unethical. The results of the survey revealed a consensus on fewer than half of the grading scenarios presented (Green et al., 2007). If assessment is a critical component both of student learning and of teachers' roles in the process, it would seem there should be greater agreement on what constitutes ethical practice. Perhaps the only element that has become clear since Starch and Elliott (1912, 1913) published their reports in the early 20th century is that teachers continue to receive little training and oversight with regards to assessment, and schools lack clear, common policies on ethical, equitable grading practices and philosophies (Feldman, 2019; Green et al., 2007).

When training is provided, it often comes at the wrong time. The Battistone et al. (2019) study noted that, in addition to the lack of training in assessment practices

for preservice teachers, those who did receive training could not apply their learnings in a timely manner so that their training became part of their practice. If teachers only receive training on assessment in their university preservice programs and not during their first year of teaching, the literature underscores their future difficulty with transferring the knowledge and skills to their classrooms (Battistone et al., 2019). Because preservice training programs vary in philosophy, focus, and requirements in terms of assessment training, graduates enter the workforce unprepared to meet the demands of fair and equitable grading (Battistone et al., 2019; Feldman, 2019). This means that professional development must take place early and continue throughout one's teaching career.

But not all professional development is equally effective in improving teaching practice (Borko et al., 2010). In 1999, researchers Stein, Smith, and Silver compared traditional professional development with what they argue is more effective teacher training. Traditional inservice that focuses on improving teachers' skills uses an outdated workshop and seminar model, while more progressive models of professional development use a variety of techniques and formats and focus on an iterative model where teachers participate in developing the agenda (Stein et al., 1999). Additional researchers have expanded on Stein et al.'s (1999) guidelines for effective professional development to include techniques such as professional learning communities and instructional coaching (Borko et al., 2010). Both new and experienced teachers need support and professional development to continuously improve, especially with regards to equitable and consistent assessment of student learning (Battistone et al., 2019; Borko et al., 2010).

Subjective Nature of Assigning Grades. In addition to a lack of training, clear policies, and oversight as to grading and assessment, the subjective nature of assigning grades creates problems. In the early years at Harvard, students were arranged according to their family's social ranking, a process that undoubtedly influenced how their work was assessed (Durm, 1993). While this might seem preposterous in contemporary academic environments, a sizable body of research indicates that grading is inherently a subjective practice that usually includes a combination of academic and non-academic evidence (Brookhart, 2013; Feldman, 2019, 2020; Guskey, 2007; Tierney, 2015). Multiple factors, such as a teacher's gender, subject area, and grade level(s) taught, appear to influence assessment decisions; similarly, students' gender, socioeconomic background, and ethnicity as well as their perceived effort appear to have influence (Tierney, 2015). While there appears to be agreement within the literature that considering gender, ethnicity, and socio-economic backgrounds is unethical when evaluating students, over 85% of participants in a 2007 teacher survey reported that they would consider effort in raising grades for lower-achieving students (Green et al., 2007). This subjective practice of considering student effort in assessing their learning may appear harmless. In fact, teachers witness their students' academic growth—growth that may be inspired by increased determination and effort—and their professional observations may sometimes yield accurate overall assessments of student learning (Guskey, 1996). However, with the potential to demotivate and disempower students due to teacher biases, the consequences of including subjective practices far outweigh the potential benefits (Brookhart, 2013; Cox, 2011; Feldman, 2019).

Implicit Bias. There is an abundance of literature pointing to teachers' implicit biases influencing student assessment, and these biases can limit student achievement and perpetuate inequities (Feldman, 2020; Guskey & Bailey, 2001; Staats, 2014). Implicit bias is a phenomenon whereby assumptions, beliefs, personal experiences, and adherence to stereotypes subconsciously influence fairness and impartiality (Feldman, 2019, 2020). There may always be an element of subjectivity in teacher grading, and research underscores that teachers typically employ teaching practices and philosophies that reflect their own personal beliefs and values (Guskey & Bailey, 2001; Guskey, 2007). In addition, research reveals that teachers employ grading practices that they themselves experienced as students and that teachers adhere to their own personal and biased philosophies of assessment whether or not these practices align with their schools' grading policies (Guskey and Bailey, 2001).

Research also reveals that even teachers committed to equitable and just grading practices are susceptible to assessing students unfairly (Staats, 2014). Some forms of implicit bias create a phenomenon known as "success bias," or the habit of believing every student can and should succeed. Success bias can inflate grades and give inaccurate information regarding student achievement. Furthermore, success bias can encourage teachers to include extraneous, nonacademic factors when determining final grades (Randall & Engelhard, 2010). This type of bias might appear harmless, but when more than 80% of the teaching population is white, students of color are disproportionately disadvantaged (Staat, 2014). In her work at the Kirwan Institute for the Study of Race and Ethnicity at Ohio State University, Research Associate Cheryl Staats (2014) outlines the problems with implicit bias, especially within predominately

White schools. She links the behavioral ramifications that result from school discipline policies that disproportionately impact students of color with consequences of grading practices that combine student behaviors, such as participation and effort, with academic achievement (Staats, 2014). Research indicates that White teachers can have lower expectations for students of color, especially Black students, and these lower expectations often are reflected in traditional grading practices. The result of implicit racial bias is that White students disproportionately enjoy the benefits of success bias, and students of color disproportionately suffer from lower expectations (Feldman, 2019, 2020; O'Connor et al., 2018).

Research shows that implicit bias influences Black students' grades to a greater degree than their non-Black peers, especially when nonacademic elements are factored into assessment (Feldman, 2019). The awarding or subtracting of points for student behaviors is inherently subjective, and when teachers and students come from different cultural backgrounds, there is great latitude for misunderstanding of intent (Feldman, 2019, 2020; Staats, 2014). Researchers have found that White teachers perceive Black students through a culturally biased lens often misinterpreting behaviors that while acceptable in the students' home life, are perceived as rude and unruly in the school environment. The literature shows this to be detrimental for several reasons. Teachers who stigmatize students in this way may cause students to modify their behavior by conforming to the negative implicit biases; may allow their own biases to modify how they teach and assess Black students; and may cause emotional responses in Black students' that negatively impact their academic outcomes (Feldman, 2019). Traditional grading practices invite implicit bias, and

while implicit bias can affect any student, research reveals it disproportionately harms students of color by lowering their grades regardless of academic performance (Feldman, 2020; Randall & Engelhard, 2010).

Assessment That Includes Non-Achievement Factors. There is considerable literature to support that academic achievement should be the primary basis for student grades (Cox, 2011; Feldman, 2019; Guskey & Bailey, 2001). Non-achievement related factors, such as including points for homework and extra credit, deducting points for late work or cheating, and rewarding effort and participation, lack concrete evidence of content proficiency and promote subjectivity and implicit bias (Battistone et al., 2019; Feldman, 2020). The results of a 2011 case study involving 16 teachers from five high schools and one adult-education program revealed that the decision to include non-achievement factors in grades was left to individual teachers (Cox, 2011). The study recorded a common refrain from teacher participants that students who turned in their homework should earn at least a C in a class; students should be rewarded for effort; and students who turned in little or no homework but earned an A on their tests should only pass the class with a C (Cox, 2011). There may be several reasons for this thinking, such as teachers have trouble distinguishing achievement from other factors and teachers consider effort and work ethic part of the habits of scholarship that students should learn in school, and therefore, upon which they should earn a grade (Brookhart, 2013; Cox, 2011).

In contrast, Feldman (2019), who writes extensively on standards-based grading practices, eschews including effort and other non-academic elements in grades. Yet similarly, Feldman (2019) considers his views on assessment a more

enlightened and progressive approach. Moreover, even though the standards-based movement emphasizes consistency and transparency, grading at the high-school level remains largely in the realm of individual teachers regardless of grading philosophy (Cox, 2011). While many schools proclaim to adhere to standards-based systems of grading, their practices often include elements of traditional grading systems thereby diluting the established goals of objective and equitable grading and doing little to address the inherent problems (Feldman, 2019). Given the body of research that indicates the extent to which traditional grading practices may promote inequitable achievement outcomes, a straightforward solution might be to switch to standards-based grading (Feldman, 2019). But schools must be willing to research, train, implement, and monitor a standards-based model for it to be effective, and even then, the methods tend to be a combination of practices that do little to alleviate the issues (Battistone et al., 2019; Feldman, 2019, 2020).

Impact of Assessment

The literature underscores a history of assessing student learning that is beset with challenges. The literature also reveals that grades and academic feedback can have real consequences for students by impacting their motivation to learn, social-emotional well-being, academic self-concept, and overall academic achievement (Feldman, 2019; Mohanty & Jena, 2015; Villeneuve et al., 2019). The impact of assessment practices for students in Catholic schools is another area of interest when examining an often-overwhelming pressure to achieve for today's students. The research on the impacts of assessment practices on students is explored below.

Impact of Assessment Practices on Students' Motivation to Learn

Researchers who study the reasons students learn found that one key element in predicting learning is students' own intrinsic motivation (Shin et al., 2017). In some students, grades may provide learning incentive by giving them the information they need to improve, and research suggests that formative feedback—both positive and negative and not necessarily in connection to a grade—plays a primary role in student motivation (Black & Wiliam, 1998; Shin et al., 2017). That said, motivation, an aspect of learning that is difficult to quantify, may provide fertile ground for teachers who employ subjective grading practices. One case study examined two groups of high school teachers from a school district with a demographic of high poverty and language diversity and reported that an overwhelming number of teachers modified their grading practices to “keep hope alive” (Cox, 2011, p. 75). In the study, the results from interviews in the first focus group revealed that teachers placed high emphasis on non-academic behaviors to motivate students, and results from the second focus group in the study confirmed the findings. However, Feldman (2019) underscores that this extrinsic motivation undermines effective teaching and learning, especially for those students who lack financial means and educational resources. Teachers confuse the muddling of non-academic factors and achievement factors in grades as proof that they care, when these actions often promote inequitable environments that perpetuate what is known as a “culture of poverty” (Feldman, 2019, p. 36). One qualitative study that sampled English teachers in grades 7-12 with at least 10 years of teaching experience examined the practice of teachers altering students' final grades and distilled the reasons to showing compassion and to potentially improving students'

future opportunities (Tierney, 2015). This completely subjective grading practice—raising one’s final grade after considering non-academic elements such as family circumstances, athletic opportunities, and the like—further muddies the waters of grading and reporting, as teachers’ grading decisions no longer even closely represent an accurate account of student achievement against the standard criteria for any given course of study (Tierney, 2015). This ultimate act of care and compassion might be viewed by some teachers and administrators as a form of social justice by assisting less advantaged students, but the literature reveals that this type of care is fundamentally unethical (Feldman, 2019) and is “imbued with power” that reinforces the racist structures found throughout American institutions (Tierney, 2015, p. 20).

One study that adds to the discourse on motivation espouses using sliding scale rubrics as a motivator for struggling students (Mahmood & Jacobo, 2019). This philosophy requires that teachers give evidence that students can succeed (or earn higher scores) through effort. Effort is assessed using a sliding-scale rubric with the underlying premise of motivating students, because the focus is on improvement over time rather than the ability to demonstrate learning at a fixed point in time (Mahmood & Jacobo, 2019). This study included 12 students who, at the onset, felt motivated by the grading-for-growth system; however, by the end, the researchers reported that only seven of the participants demonstrated noticeable growth, and five of the participants demonstrated no growth at all. While the sample size in this study is small and the researchers reported the results were inconclusive, they committed to further exploration of more equitable grading systems based on these initial findings (Mahmood & Jacobo, 2019).

Accurate assessment information allows for student self-reflection, a valuable component of long-term, sustained student motivation to learn (Brookhart, 2008). The literature supports the positive effects of meaningful, clear, earned, and appropriate feedback, but grades by themselves should not be confused with this type of feedback as a means to motivate (Brookhart, 2008; Feldman, 2019). Black and Wiliam (1998) assert that grades are often overemphasized in schools—especially when they resemble a system of “gold stars”—while useful advice in the form of feedback is underemphasized (p. 143). Practices developed to enhance student motivation, including visible or public systems of rewards, do little to inspire intrinsic motivation and conversely produce negative effects. Comparing students by class ranking or the use of a curve sends the message that student achievement is meaningful only if it is higher than other students’ achievement (Feldman, 2019; Shin et al., 2017). While this practice may do little to motivate any student, it can do harm to at-risk students. In fact, when struggling students are evaluated in comparison to other students, it negatively impacts the motivation to achieve, increases anxiety, and can even promote a withdrawal from learning altogether (Black & Wiliam, 1998; Brookhart, 2008; Shin et al., 2017).

Impact of Assessment Practices on Students’ Social-Emotional Health

For students, grades can mean much more than just the determining factor in course placement, athletic eligibility, college admission, scholarship opportunities, and future employment options. Grades can be the source of crippling stress and anxiety and can wreak havoc with their self-esteem (Feldman, 2019; Mohanty & Jena, 2015; Villeneuve et al., 2019). The relentless pressure on high-school students to achieve

often comes at the expense of sleep, building resilience reserves, and physical and emotional health (Feldman, 2019). Research overwhelmingly shows that grades are a major source of anxiety and stress for students, and many common grading practices amplify this stress (Feldman, 2019; Villeneuve et al., 2019).

Rates of adolescent anxiety are on the rise, and students in both public and private schools report that school-related stress is the primary contributor (Pew Research Center, 2019). This may be more pronounced for students who attend high-achieving high schools who report that they willingly compromise learning and their own health and well-being for the sake of high academic outcomes (Twenge et al., 2019; Villeneuve et al., 2019). To assist schools in implementing programs to improve student well-being, Villeneuve et al. (2019) surveyed approximately 175,000 middle- and secondary-school students from 54 schools asking questions related to homework, physical health, and school-related stress. Close to 40% of the high-school students reported missing school at least one time in the past month for a health or emotional problem, 70% reported experiencing exhaustion, and between 68% - 88% reported being often or always stressed by their schoolwork load. The results also revealed that almost 80% of the participants reported grades and other assessments as the reason for their stress, a problem that often can be overlooked when these schools are viewed as exemplars due to their students' academic performance (Villeneuve et al., 2019). Perhaps even more concerning with regards to student wellness is that 30% of the high-school students surveyed reported having little or no confidence when it came to their abilities in coping with their feelings of stress (Villeneuve et al., 2019).

Impact of Assessment Practices in Catholic High Schools

According to a report by Peter Imperial (2012), an educational scholar who writes extensively about grading in Catholic schools, Catholic schools are similar to secular schools in that students report feeling pressure to earn good grades, and teachers employ problematic grading practices, practices that leave both students and teachers feeling frustrated (Imperial, 2012). Like secular schools, many Catholic high schools throughout the country identify with the pressures of high-achieving, college-preparatory environments. Imperial (2012) describes the difficulties for Catholic high schools in alleviating the causes of stress and anxiety due to the philosophical tension between the mission of serving student populations who need academic support and the confusion that comes with holding less-advantaged students to the same account as their more-advantaged peers (Imperial, 2012). Imperial (2012) notes how Catholic schools are not immune to the pitfalls often associated with many grading practices that include non-achievement criteria when determining final grades; he also notes the lack of research that explores grading practices in Catholic schools. Because of this gap, Imperial conducted his own study in 2012 that surveyed 428 Catholic high-school teachers. Imperial's (2012) research revealed that 78% of the teacher participants included homework scores in final grades; 57% included effort; and over 30% included neatness. In other words, the Catholic-school teachers in this study combined non-academic factors with academic-achievement factors to arrive at final grades. As private schools, Catholic schools are neither uniform in governance nor assessment philosophy, and Catholic teachers often operate with less oversight and greater latitude than their public-school peers (Bryk & Holland, 1993; Imperial, 2012). Given the

research on the need for on-going professional development on equitable grading practices as well as the impact of common grading practices on student learning and social-emotional well-being, the absence of research into grading practices at Catholic high schools potentially warrants more focus and resources. Catholic-school mission statements that often include phrases such as “educating the whole child,”—a philosophy that equates moral development and social-emotional health with academic achievement—suggests further exploration on best practices is required (Imperial, 2012; McDermott, 1997).

Teacher Feedback

The literature identifies that feedback, the formative and evaluative information given to students that describes their academic performance, as a critical component in the learning process, yet not all feedback is equally effective (Hattie & Timperley, 2007; Poulos & Mahony, 2008). In a synthesis of 12 meta-analyses focused on classroom feedback that included 196 studies encompassing various grade levels and school settings, feedback fell in the top five to 10 most important influences on student achievement (Hattie & Timperley, 2007). Likewise, a study that surveyed 146 ninth graders (aged 14 to 17 years) reported on the benefits of feedback and noted that students identified perceiving process-oriented feedback as more useful than grade-oriented feedback (Harks et al., 2014). Researchers Hattie and Timperley (2007) and Harks et al., (2014) reported that the most valuable forms of feedback provided reinforcement or learning cues. But for feedback to be effective, it also must be delivered in a manner that accounts for students’ prior knowledge and their overall achievement level (Shute, 2008). Put differently, students who struggle may require

that formative feedback be delivered in a different manner from their more proficient peers (Shute, 2008). Therefore, there is no one best type of feedback; effective feedback may take on different formats and delivery styles to promote student learning (Shute, 2008).

Yet the literature suggests that there are forms of feedback that are less effective, such as feedback that offers overly general praise and punishment (Harks et al., 2014; Hattie & Timperley, 2007). The studies reveal that comments such as suggesting students simply work harder or study more to improve their grades have little or no impact, while specific language not only reduces the likelihood that feedback is oversimplified, it also provides transparency on improvement strategies, allows for student self-reflection, and gives agency toward improving achievement (Hattie & Timperley, 2007). Feedback that takes the form of extrinsic rewards, such as stickers, prizes, and awards, significantly undermine students' intrinsic motivation to learn for much the same reason as publicizing grades (Feldman, 2019). These types of tangible rewards increase competition among students and communicate that being the best is more important than learning a concept, strategies that have not been shown to be effective (Feldman, 2019; Hattie & Timperley, 2007; Shin et al., 2017). While determining best feedback and grading practices is important for all students, research suggests that females experience this type of feedback differently compared to their male counterparts, and therefore, their experiences require additional learning and consideration (Johnson & Helgeson, 2002).

Females' Experiences with Assessment and Feedback

Understanding the impact of how females experience grades and feedback might highlight the importance of evaluating common high school grading practices (Johnson & Helgeson, 2002). In Ryan and Henderson's (2018) study of 4,514 university students, students were more likely to reject teachers' evaluative feedback if the comments evoked negative emotional responses. Their study found that for students to be open to receiving feedback and willing to act upon it, they need to be agreeable to the message (Ryan & Henderson, 2018). Research reveals that this is especially true for female students, and research suggests that females of all ages interpret and react to feedback differently (Altermatt & Pomerantz, 2003; Johnson & Helgeson, 2002). Altermatt and Pomerantz (2003) examined a sample of 932 elementary students and found that girls admitted to worrying about their grades, pleasing the teacher, and receiving negative feedback on their academic performance to a far greater degree than boys. For example, females tend to agree with feedback and make academic behavioral changes based on the feedback they receive to a greater extent than their male counterparts (Johnson & Helgeson, 2002). Literature also reveals that female students, no matter if they are middle-school, high-school, or college-aged, process academic experiences differently from males (Altermatt, 2015; Johnson & Helgeson, 2002; Pomerantz et al., 2002). Whether the experiences are perceived as positive, such as receiving a good grade on an assessment or positive feedback on an assignment, or negative, such as receiving a poor grade or answering a question incorrectly, girls reported experiencing more total interactions with the teacher overall (Altermatt, 2015). Literature suggests that one possibility for this

difference is that girls take achievement-related feedback more personally and are likely to register more experiences or encode neutral experiences as negative to a greater degree than boys (Altermatt, 2015). Gender research supports this notion showing that females and males exhibit different patterns of processing and responding to feedback in achievement situations (Roberts, 1991).

The literature identifies inequities in assessment practices due to teacher biases that are rooted in gender discrimination. Mechtenbeg (2009), to better understand gender differences related to the phenomenon that most highly educated women earn less than their male counterparts, reviewed studies that examined the behavior of teachers and students in both math and humanities classrooms. The review found that well-intentioned teachers, believing they are battling gender stereotypes, praised female students in the math classroom, and the praise produced the opposite of the intended effect (Mectenberg, 2009). Believing the praise to be false, female students develop a mistrust, and the study posits that this type of biased feedback has consequences that reach beyond the classroom, impacting girls' performance in math and science, and discouraging girls from choosing math and science majors that allow for high-earning jobs later in life (Metchenberg, 2009).

Evidence reveals girls typically outperform boys academically, though there is a growing body of evidence showing that girls are more prone to internal distress associated with grades and evaluative feedback, distress that increases self-doubt and negatively impacts self-esteem (Pomerantz et al., 2002). Girls also link their academic success to their innate academic abilities and overall intellectual capacity (or lack

thereof), and both positive and negative academic experiences impact their motivation, self-efficacy, and self-esteem to a greater degree (Altermatt & Pomerantz, 2003).

Impact on Female Self-Esteem

The focus on girls' experiences in school began in the 1970s when the possibility that girls experience internal distress because of academic performance came to light (Pomerantz et al., 2002). This precipitated a body of research that focused on the reasons that girls experience academic stress and the potential negative outcomes that result (Altermatt & Pomerantz, 2003). While the American Psychological Association (2003) reports that the presence of stress and anxiety is normal in all adolescents, a number of studies show that females are impacted to a higher degree than males (Altermatt & Pomerantz, 2003; Jena and Mohanty, 2015). In fact, a meta-analysis of 26 studies involving approximately 60,000 observations of children born between 1965 and 1996 revealed that adolescent girls experienced significantly higher rates of anxiety and depression (Costello et al., 2006). These findings suggest that girls are more concerned with pleasing adults, which would account for their experiencing internal distress when they perceive that their performance has not pleased a teacher (Pomerantz et al., 2002). While this internalization may work to intensify their motivation to improve, girls run the risk of setting unreasonable expectations for themselves resulting in perfectionism, internal distress, depression, and anxiety (Madigan, 2019; Pomerantz et al., 2002).

Impact on Female Academic Motivation and Self-Efficacy

The literature indicates that girls exhibit decreased motivation to achieve when they have been previously unsuccessful, and that they tend to blame poor performance

on their innate lack of ability (Roberts, 1991). One possible explanation for this is that girls show a greater dependency on external evaluations when assessing their own performance (Roberts, 1991). One early study by Rosenberg and Simmons (1975) revealed that girls were concerned with others' evaluations of their performance to a greater degree than boys, but this does not necessarily mean that females evaluate themselves more negatively, only that their self-evaluations can be influenced by others' feedback to a greater degree (Roberts, 1991). There is considerable literature that shows that while girls generally perform better than boys in school and they graduate with higher grade point averages, their performance on standardized tests does not match this success (Bian et al., 2018; Duckworth & Seligman, 2006; Pomerantz et al., 2002). One possible explanation for this is that teachers reward students for non-achievement factors, and girls are more self-disciplined and more willing to engage in behaviors that influence teachers' perceptions of their academic abilities (Duckworth & Selgman, 2006; Feldman, 2019). Yet, grading on non-achievement factors differs from offering specific feedback on these factors as it no longer is formative in nature, and teachers run the risk employing common biased and inequitable grading practices—practices that in the long-term negatively impact motivation and self-efficacy (Altermatt & Pomerantz, 2003; Feldman, 2019).

Research shows that there is a link between girls' academic achievement and their motivation while in school and beyond (Alivernini & Lucidi, 2011). One factor that negatively impacts motivation is when girls perceive feedback to be unjust or negatively gender stereotypical (Rinfret et al., 2014). In this case, girls may react by psychologically disengaging from the feedback and retreating from their studies. This

form of self-preservation negatively impacts the learning process by attributing no legitimacy to the feedback and devaluing school and the role of education overall (Major & Schmader, 1998; Rinfret et al., 2014). A 2014 study by Rinfret et al. involving 236 female high school students reported that a connection exists between receiving poor grades and female students' motivation to succeed, especially when they perceived their teachers were not fairly assessing their performance. Moreover, the students reported that the evaluative feedback from these teachers decreased their motivation to continue their studies in a particular content area, a result that may have consequences related to overall school satisfaction, as well as implications for future career and earning opportunities (Rinfret et al., 2014).

There is research that points to overall bias against girls' intellectual ability in academically competitive environments; therefore, girls responding negatively to feedback they perceive as gender biased is an expected outcome (Bian et al., 2018). A 2015 Pew Research Center poll showed that 86% of participants ascribed the term "intelligent" equally to both men and women. At the same time, college students use the terms "brilliant" and "genius" two to three times more often when evaluating their male professors (Storage et al., 2016). Google searches reveal inquiries about whether sons are gifted 2.5 times more often than whether daughters are gifted (Stephens-Davidowitz, 2014). Girls may experience stereotype-relevant feedback differently because they belong to the group in which the stereotype exists (Biernat & Danaher, 2011). In a study designed to examine how members of a stereotyped group translate feedback, female students interpreted subjective language, like "good," "great job," or "not bad" to have a negative connotation (Biernat & Danaher, 2011). And when

feedback—even positive feedback—included stereotypical words such as “emotional,” female students interpreted it to indicate lesser praise than did male students (Biernat & Danaher, 2011). Like the findings in the Rinfret et al. study (2014), the consequences of this type of feedback included lower quality of post-feedback performance, decreased motivation to invest in future academic performance, and decreased academic self-efficacy (Biernat & Danaher, 2011). These findings are important when considering how girls experience feedback because research suggests that students who perceive themselves as more academically capable will achieve better academic results (Bian et al., 2018; Johnson & Helgeson, 2002).

Summary

There is much research to support that common grading practices in American schools are laden with inconsistencies, and educational researchers have raised concerns regarding the impacts of these practices on students. Remnants of traditional grading conceived over a century ago are muddled with more current standards-based practices popularized in the 1980s to form the *mélange* present in schools today. While feedback in educational contexts is considered crucial to the learning process, motivation, and overall achievement, teachers are unclear on what types of feedback are most effective (Altermatt & Pomerantz, 2002; Feldman, 2019). These practices frustrate teachers and students—teachers report receiving little training and oversight, while students report experiencing increased school-based stress and anxiety as a result of unclear and inequitable practices. There is also a significant body of research about how students respond to assessment and evaluative feedback, but there is less research focused specifically on female students’ experiences. That said, the scant

literature specifically addressing how girls perceive feedback underscores that girls experience feedback differently and describes the inherent risks that non-specific, stereotypical, and negative feedback has on girls' self-efficacy, motivation, emotional health, and overall well-being. More is needed to fill this gap. To create equitable experiences for girls, it is important to understand current grading practices and to study how girls experience feedback and assessment. This study endeavors to fill the gap on the impact of feedback and assessment practices at an all-girls, urban, faith-based school in order to identify areas of improvement that will better equip teachers to serve the female students in their care.

Chapter 3: Methodology

The following chapter outlines the methodology used in this mixed-methods case study to investigate the impact of feedback and assessment practices on female students in an all-girls high school in an urban, faith-based environment. It reviews the research questions and describes the rationale, methods used for selecting the research participants, and steps taken for collecting and analyzing the data.

Research Questions

The purpose of this case study is to understand the impact of feedback and assessment practices on female students in an all-girls high school in an urban, faith-based environment. The specific research questions guiding this study were:

1. What do teachers at an all-girls, faith-based urban high school perceive to be formative and summative assessment strategies that support improved student learning in their classroom?
2. How do female students experience feedback and assessment practices in an all-girls, faith-based, urban high school?

Rationale for Methodology

A mixed-methods case study design was used to address the research questions in this study. The case study is appropriate for this research as it focuses on complex functioning within a natural context, is bounded by parameters, such as time and place, and explores a real-life situation through multiple sources of data collection, in this case both quantitative surveys and qualitative open-ended interview questions (Creswell & Poth, 2018; Yin, 2018). This case study was bounded by several criteria, specifically it was centered within one all-girls, faith-based, urban high school. In

addition, the bounded system included two units of analysis: faculty and students. According to Yin (2018), case study research addresses “how” and “why” questions, and because this study focuses on one school and explores how students experience feedback and assessment, the case study is an appropriate approach. Yin (2018) advocates for both quantitative and qualitative data-collecting methods in case study research, which fits with this study’s use of surveys and follow-up semi-structured interviews to collect data (Yazan, 2015; Yin, 2018).

Specifically, an explanatory sequential design was used, which begins with quantitative data collection and analysis followed by qualitative data collection and analysis to further expand on the quantitative findings (Gay et al., 2012). Case study research often leans toward a realist perspective, or a single reality, whereas this case study is oriented toward a relativist perspective, where a constructivist approach was used to capture the perspectives of a number of student participants to determine any common themes in their experiences (Yin, 2018).

Context

Given the unique nature of the context for this study, the state and location will remain undisclosed to protect the anonymity of the school and participants. The school in this study refers to itself as an all-girls school. Since this study was conducted in 2021-2022, when there are many meanings to the word “girl,” discussion of the term is necessary. All-girls or all-boys schools are commonly referred to as single-sex schools, although this can be confusing as not all single-sex schools share common definitions or requirements when considering what makes them all-boys or all-girls. There is some consensus in the literature that the term sex relates to biology and the

two sexes, male and female, and the term gender relates to behavior, aptitudes, and appearance based on sex, societal expectations, and cultural influences (Lau et al., 2020; Mazure, 2021). *Sex* is assigned at birth, while *gender* is how a person identifies. The idea that there are only two genders is sometimes referred to as “gender binary,” and the idea that gender identities are neither male nor female is sometimes referred to as “non-binary” (Lau et al., 2020). The participating school’s current policy (2022) dictates that students who identify as girls and individuals who were assigned female at birth and identify as non-binary may be considered for admission. The school further refines this policy and states that students need to feel they belong in a community of women and will support the institutional language that privileges the female voice. Throughout this study, the terms “girls,” “females,” and “young women” are used interchangeably to describe participating students and students referred to throughout the literature. This imprecise usage is difficult to avoid when considering the various definitions and settings throughout the literature. While all participating students in this study attend one school that identifies itself as an all-girls school, there is no school policy requiring the students to provide official documentation, such as a birth certificate, to identify themselves, and they were not asked to identify themselves in this study. In addition to the discussion on student gender identification, the term “girl” may be used for older students and is not meant to be demeaning or pejorative.

Description of Setting

This study was set in an all-girls, faith-based, college preparatory, urban high school in the northwestern United States founded in the mid-19th century. The total

enrollment for a typical year ranges from approximately 650-700 students, and most students attend an institution of higher learning after graduation. The annual tuition is approximately \$16,000, and between 37% to 50% of students receive tuition assistance, with an average financial aid package of approximately \$8,000. The student body is predominately White at 68%, with 10% multiracial students, 6% Hispanic students, 6% Asian/Pacific Islander students, and 5% Black students. The average class size is 21 students, and the student-teacher ratio is 13:1. The school typically employs a faculty and staff size of 90 to 100 individuals, and 88% of the 56 faculty members hold advanced degrees.

Participants

To determine how female students experience feedback and assessment in an all-girls, faith-based, urban school, data were collected from two participant groups, or two embedded units of analysis—faculty and students—using criteria purposive sampling and intensity sampling (Yin, 2018).

Faculty Participants. Faculty participants were surveyed using a purposive sampling strategy that allows for the selection of participants based on their ability to address the research questions to develop a better understanding of the research problem (Creswell & Poth, 2018). I chose this strategy over other sampling strategies to obtain a range of experience among faculty and faculty of different genders and ethnicities. The criterion for inclusion consists of being a faculty member present at the initial faculty inservice yielding a potential maximum sample size of 56 participants. Typically, attendance at inservice meetings ranges from 95% to 100% of faculty members present.

Student Survey Participants. Student participants were surveyed using purposive sampling, a sampling strategy that in quantitative research allows (a) an intentional selection of sample participants; and (b) the best opportunity to access a range of experiences within a chosen population (Gay et al., 2012). The student survey was administered to all juniors and seniors who expressed interest in participating in the study and who received parent permission to do so. Of the 340 juniors and seniors currently enrolled, 101 students (56 juniors, 45 seniors) met those criteria. Juniors and seniors were selected because freshmen students would have had little experience with current feedback and assessment practices as they were new to the school, and sophomore students would have had little experience due to the COVID-19 pandemic interruption. While juniors and seniors experienced a significant interruption as well, they were more likely to have experienced current feedback and assessment practices at the school during their freshman and sophomore years.

Student Interview Participants. Student interview participants were selected using purposive sampling. Students were identified for Phase 2 of this study from an email that introduced the survey and stated that if students were interested in participating in a follow-up interview, they should respond to the email for consideration. From the respondents who indicated a willingness to be interviewed, a purposive sample was drawn that represented a balance of grade level (junior and senior). The target sample size for Phase 2 was determined by the responses in Phase 1.

Design and Procedures

The explanatory sequential design of this case study included multiple methods, and data were collected in two phases, which included one faculty survey, one student survey, and semi-structured student interviews. Because the research purpose brought together a closed-ended quantitative database (surveys) and an open-ended qualitative database (follow-up interviews), the mixed-methods approach was suitable (Creswell & Poth, 2018; Gay et al., 2012). Using the survey instruments, I collected demographic data on both units of analysis (faculty and students), as well as information from faculty members on their perceived skill with formative and summative assessment strategies that support improved student learning, and information from students on how they experienced grades and academic feedback. Because of my role as principal of the school, I worked with two neutral third parties, one for students and one for faculty, to collect the data in both phases of the study.

Phase 1

During Phase 1 of the study, two instruments were used. The first instrument was a survey that a neutral third party, a doctoral fellow at the University of Portland, introduced in person to all faculty present during a faculty inservice and sent electronically during the meeting for participants to complete using Qualtrics software. Faculty members were emailed a meeting agenda a few days in advance with a reminder to bring their laptops, a request that follows typical meeting protocols. The agenda included “Faculty Survey on Assessment Practices,” so the faculty was aware of the survey in advance. After IRB approval was granted for the study, the neutral third party conducted the survey during a September 2021 faculty inservice.

The second instrument in Phase 1 is a student survey that junior and senior students completed after expressing interest in participating in the study. Students are often asked to give feedback to the school via electronic survey, but as the information is being collected in my role as doctoral student and not principal of the school, parents were asked in advanced to grant permission for their student's participation in the survey. To determine the student participants, the school librarian emailed all juniors and seniors describing the study and giving information on how to participate. Juniors and seniors who expressed interest were then cross-checked with juniors and seniors who received parent permission. Those students were given instructions on how and when to participate in the survey. The school librarian conducted the survey in October 2021 during a school "flex" block, a period during the day where students participate in a variety of activities. Participating students brought their school-issued iPads and met in the auditorium to take the survey, and all were finished within 20 minutes.

Phase 2

Phase 2 of this study used semi-structured interviews with survey participants from Phase 1 who responded to an email expressing interest in participating in the study's follow-up interviews. This phase of the study allowed for the collection of valuable qualitative data about how students experience feedback and assessment. A neutral third party, the school librarian, identified a sample that included both juniors and seniors. Initially, the goal was to represent the diversity of the school population within the sample, and three of the 10 interview participants ($n = 30\%$) identified as students of color. To learn more about how female students experience academic

feedback and assessment, the librarian contacted the participants, scheduled interviews, interviewed the participants on Zoom, and prepared the transcripts. Table 1 describes the timeline of the procedures for this study.

Table 1

Timeline of Procedures

Dates	Procedures
Aug. 2021	Proposal defense, permission secured for instrument use and adaptation, IRB approval completed
Early Sept. 2021	Met with third-party survey administrator/interviewer to communicate process and schedule dates for Phase 1 surveys
Sept.-Oct. 2021	Administered surveys
Early Oct. 2021	Analyzed data from Phase 1
Mid-Oct. 2021	Third party identified participants and scheduled interviews
Late Oct.- Early Nov. 2021	Third party conducted and transcribed interviews
Mid-Nov. - Mid-Dec. 2021	Coded data
Dec. 2021 – Feb. 2022	Completed data analysis; wrote Chapters 4 & 5

Table 2 outlines the research questions, corresponding phases of the study, and the instruments used during each phase.

Table 2*Instrument alignment with phases of the study*

Phase	Research Questions	Instrument
Phase 1	What do teachers at an all-girls, faith-based urban high school perceive to be formative and summative assessment strategies that support improved student learning in their classroom?	Faculty Survey: Scale of Teacher Assessment Practices
Phase 1	How do female students experience feedback and assessment practices in an all-girls, faith-based, urban high school?	Student Survey: Instructional Feedback Orientation Scale
Phase 2	How do female students experience feedback and assessment practices in an all-girls, faith-based, urban high school?	Semi-Structured Interview Protocol

Instruments

The faculty and student surveys allowed for the collection of valuable quantitative data, and the follow-up semi-structured interviews provided a more complete picture of the study. The use of multiple instruments in a case study allows for triangulation, which helps strengthen the case and ensure trustworthiness (Gay et al., 2012).

Phase 1: Faculty Survey Instrument

To address Research Question #1, the Scale of Teacher Assessment Practices (STAP) instrument was used (Appendix A). The STAP was adapted from the Assessment Practices Instrument Revised (API_R) (Burry-Stock & Frazier, 2005), an instrument designed to assess seven critical competencies essential to a teacher's role

as determined by the “Standards for Teacher Competence in Educational Assessment of Students” (1990). This sets forth those teachers need the following skills:

1. Choosing assessment methods appropriate for instructional decisions
2. Developing assessment methods appropriate for instructional decisions
3. Administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods
4. Using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement
5. Developing valid pupil grading procedures which use pupil assessments
6. Communicating assessment results to students, parents, other lay audiences, and other educators
7. Recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information. (American Federation of Teachers et al., 1990, pp. 4-6)

To develop the instrument, the STAP was administered to 193 elementary school teachers in nine elementary schools in the southeastern United States. The questions were selected and modified from the API_R and were designed to measure six domains of assessment literacy. These domains were distilled from the seven domains adopted by the American Federation of Teachers, and the characteristics of the six domains in the STAP contain all the elements of the seven domains, but the wording is modified for concision. The instrument asked participants to indicate their skill level with regards to assessment practices, using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The developer of this

instrument acknowledges deviation from the traditional Likert scale and gives the rationale that the middle number on this adapted scale represents a “skill level” and not the choice to “opt out” (Howell, 2013). Table 3 shows the six domains from Howell’s (2013) STAP instrument (p. 26) and the corresponding prompts.

Table 3

Research Question #1 - Faculty Survey: STAP Assessment Areas by Domain and Question

STAP Assessment Domains	STAP Criteria
1. Selection and development of assessment methods	3, 11, 15, 17, 18, 27, 29
2. Administering, scoring, and interpreting assessment results	2, 13, 14, 19, 25, 26
3. Using assessment results to inform day-to-day decisions	4, 6, 9, 22, 23, 24
4. Communicating assessment results to others	1, 7, 8, 16, 21, 30
5. Ethical use of assessment	5, 10, 12, 20, 28
6. Overall skill with assessment	31, 32, 33, 34, 35, 36

Below are some sample assessment criteria from the STAP instrument in which teachers rated their skill using the 1-5 rating scale:

- Using the results of formative assessment to adjust the content of lessons
- Using results of summative assessments to adjust future lesson plans
- Creating assessments that accommodate the needs of a variety of students

- Communicating the results of assessments to students in a way that they can understand
- Explaining to students how assessment results will be used to assign grades

To determine the reliability of the STAP instrument, Cronbach's alpha was calculated for all items and was .96 ($M = 117.46$; $SD = 16.42$) indicating strong internal consistency reliability (Howell, 2013).

Phase 1: Student Survey Instrument

Before the date scheduled for the student survey, a neutral third party, the school librarian, sent participating students an email letting them know how and when the survey would be conducted. The survey took approximately 10-20 minutes to complete, and participants used their school-issued iPad to participate in the survey. Students who expressed interest in participating and received parent permission, but who were absent on the day the survey was conducted were not included in the survey. Because of the close-knit community at the participating school, it is believed that students would have a difficult time establishing trust with an unfamiliar adult. So a neutral third party, the school librarian, was selected to administer the survey and conduct the interviews because of her considerable experience with action research, conducting interviews, and the culture of the school setting and population. The librarian is a known and trusted adult within the student community, and she has no direct influence on students' grades.

The survey instrument was a 27-question questionnaire that used a five-point Likert-type scale and assessed the perceptions of instructional feedback (King et al., 2009). This questionnaire, the Instructional Feedback Orientation Scale (IFOS)

(Appendix B), is the culmination of researchers Paul King, Paul Schrodt and Jessica Weisel's efforts to find a reliable and valid measurement of how students experience academic feedback. Their work demonstrated that there is a gap between students' observed performance and the actual desired performance, a discrepancy, or "feedback-standard gap," which led the researchers to develop a tool to measure students' responses to feedback (King et al., 2009, p. 237). The IFOS originally contained 124 questions and went through two iterations before its final format. In the first study, researchers, using a sample size of 212 college students from both a community college and large private university, developed and piloted the instrument they designed to assess students' experiences with teacher feedback. Students took approximately 35 minutes to complete that 124-item questionnaire. A major concern with this study was reliability due to the number of items and the onset of survey fatigue (King et al., 2009). The results from this first study led the researchers to develop the IFOS, a pared-down, 27-item questionnaire that addresses four dimensions of students' responses to teacher feedback, namely utility, sensitivity, confidentiality, and retention (King et al., 2009). The testing and iterations of the instrument resulted in "estimates of internal reliability for the IFOS produced acceptable Cronbach's alpha coefficients for each dimension, ranging from 0.69 for feedback retention and 0.74 for feedback confidentiality, to 0.86 and 0.88 for feedback sensitivity and feedback utility respectively" (King et al., 2009, p. 244). Another study the researchers conducted sought to establish validity for the IFOS. The researchers determined that there was evidence of concurrent and discriminant validity, and the "magnitude of the correlations between the IFOS dimensions and other student

characteristics measured. . . provide further assurance that the IFOS” is a useful instrument in the assessment of students’ experiences with instructional feedback (King et al., 2009, p. 255). The researchers emphasize the need for further exploration into how students experience feedback as these data are an essential tool in improving the practice of teaching and learning. The survey concludes with a demographic section. This section includes a question about race/ethnicity that replicates the race/ethnicity categories the school uses to gather enrollment data.

Phase 2: Student Semi-Structured Interviews

The interview protocol (Appendix C) used in Phase 2 is designed to better understand how female students experience academic feedback and assessment. Interviews were conducted with a random sub-sample of student participants who gave permission to be contacted for a follow-up interview on a separate email that introduced the survey. The process of creating interview questions, establishing rapport between interviewer and interviewee, conducting the interviews, and transcribing the results are challenges within any case study; therefore, it was important to choose and train the third-party interviewer carefully (Creswell & Poth, 2018). I met with the librarian for protocol review before the interviews.

Student participants were notified in advance that they would be asked questions about feedback on both a specific assignment and as well as feedback in general. As part of the interview process, they were asked to bring an assignment, project, lab, or test that had teacher feedback (Appendix D). Interviews were approximately 13-15 minutes each and semi-structured, which allowed for the interviewer to start with predetermined questions about the artifact and students’

experiences with the feedback, as well as to give opportunities for follow-up questions to clarify or expand answers (Wilkinson & Birmingham, 2003). The third party conducted all interviews via Zoom video conference and recorded each interview.

Before the start of the interview, the interviewer established rapport to help correct for the inherent power differential between interviewer and interviewee (Creswell & Poth, 2018). The interviewer has established herself within the school community as friendly, welcoming, and positive. Depending on whether she and the interviewee had existing rapport, she began the interview with a question like: “How are you doing today” or “How is your year going so far?” Once she determined the participant felt comfortable, she explained the interview process and the need for honest responses with as much detail as possible. She assured students that their responses will be coded for confidentiality. The question design in the interview protocol was informed by:

1. analysis from the student survey using the Instructional Feedback Orientation Scale (King et al., 2009),
2. themes found in the literature on how female students experience academic feedback and assessment, and
3. the research question that explores how female students experience feedback and assessment practices in an all-girls, faith-based, urban high school.

The seven interview questions were peer reviewed by five doctoral students and two professors from the participating school to strengthen the clarity and reliability of the questions (Creswell & Poth, 2018). Before the study, questions were asked to two students, who would not be participating in the study, to investigate how

students would respond to the questions. This helped refine the interview before the training session with the third-party interviewer. Some sample interview questions included:

- Please describe the assignment you brought with you.
- Talk with me about why you chose this assignment/test.
- Thinking about assessments or feedback you have received from teachers at your school, can you describe any that stick out as being particularly encouraging or discouraging and explain why?

The third-party interviewer transcribed the interviews and provided copies of the Zoom recordings and interview transcripts. Along with reviewing the transcripts, I viewed each recorded interview and wrote analytic memos describing body language and non-verbal communication.

Role of the Researcher

The researcher should be adaptive, flexible, and unbiased (Yin, 2018). As a former English teacher, I experienced the uncertainty and frustration associated with assessing student work and providing effective feedback to help improve student learning. The constant pressure of keeping up with a full teaching load, in this case five English courses with three distinct course preps, left little time to norm practices and hone skills in providing feedback. During my teacher preparation courses and 13 years of experience in an all-girls, faith-based setting, I received little preservice training or professional development about how to give effective feedback, equitable assessment practices, or feedback and assessment practices specific to girls. I also experienced a wide variety of student reactions to feedback from both formative and

summative assessments. The reactions ranged from gratitude for assistance in helping the students improve to negative reactions that, in some cases, led to a breakdown in the teacher-student relationship. Because these experiences may lead to personal bias, I continually took steps in my data collection and analysis to address potential bias. Because my current role was principal, I was in a supervisory and evaluative role regarding the teachers involved in this study, and I was the school's senior administrative official regarding all students. For these reasons, it was necessary to use a neutral third party to administer surveys and conduct follow-up interviews. In this way, the teachers under my supervision would feel more comfortable speaking freely about their experiences, and students would not fear any potential repercussions from teachers or administrators by their honest participation in the study.

Data Analysis

The quantitative and qualitative data were collected from two unique sources and in two ways to better understand the research questions, which allowed for triangulation and increased trustworthiness of the study (Creswell & Poth, 2018; Yin, 2018).

Quantitative Analysis

Descriptive statistics were used to analyze the quantitative data to identify patterns. Each survey instrument was organized into themed groupings. The faculty questionnaire (STAP), a tool that seeks to determine whether patterns emerge with faculty's self-perceived skill with various assessment practices, groups the questions into these six categories:

- selection and development of assessment methods,

- administering, scoring, and interpreting results,
- using results to inform day-to-day decisions,
- communication of results to others,
- ethical use of assessment, and
- overall skill with assessment.

Once the demographic information was disaggregated, patterns emerged between the teachers' demographic categories and their perception of skill interval range. The interval data were analyzed using independent sample *t*-tests and one-way ANOVAs to compare by demographic group depending on group sample size. The student questionnaire (IFOS) was also pregrouped into categories, namely utility, sensitivity, confidentiality, and retention. Patterns emerged by category that gave insight into the different factors involved in how students experienced academic feedback. All quantitative data were analyzed using Excel. Findings are shared in Chapter 4 and analysis, along with qualitative data analysis, considered in Chapter 5.

Qualitative Analysis

The qualitative open-ended student interview responses were analyzed in cycles. During the first-cycle coding, I reviewed the transcripts and assigned headings that describe the data (Creswell & Poth, 2018; Saldaña 2016). Provisional a priori codes included terms such as *motivation to learn*, *academic self-efficacy*, *social-emotional well-being*, and others that appeared repeatedly during the analysis. During the second-cycle coding, I grouped the headings into categories, and compacted the number of categories with appropriate warrants (Saldaña, 2016). Themes that emerged are presented in Chapter 4 and are analyzed and interpreted in Chapter 5.

Connecting the Data

According to Gay et al. (2012), in an explanatory sequential design, the final step is to connect the findings from the qualitative data back to the quantitative data in the first phase of data collection. In this study, the quantitative and qualitative data considered together provided a more in-depth understanding of the study's research questions.

Ethical Considerations

The research was conducted with the highest regard to ethical considerations and minimizing potential risk for the participants and the school. Data were collected upon approval by the Institutional Review Board (IRB) at the University of Portland during the fall of the 2021-22 school year. I took deliberate steps to ensure confidentiality for all faculty participants and student participants. Participant information was coded to protect their identity. All data were stored electronically on my password-protected Dropbox account and were copied to my password-protected external hard drive. All participant codes were stored on a secure device separate from the raw data containing those codes.

Ensuring Quality

To ensure the quality of this study throughout the design, data collection, and data-analysis phases, I adhered to and provided evidence of trustworthiness including elements of credibility and reliability, dependability, transferability, and confirmability (Lincoln & Guba, 1982). To further establish the trustworthiness of this single case study with two embedded units of analysis (Yin, 2018), I followed Lincoln and Guba's

(1982) traditional criteria, which includes truth value, applicability and consistency, and neutrality.

Truth Value

I established confidence in the veracity of the data collected by providing information on the participants, setting, design of the study, and methods used to ensure confidentiality. By corroborating evidence from multiple sources, in this case study through faculty surveys, student surveys, and student interviews, I could ensure quality and trustworthiness through triangulation (Creswell & Poth, 2018).

Applicability and Consistency

I detailed the participants and setting, in this case female participants in an urban, all-girls, faith-based high school, and the methods of data collection and analysis. It can be assumed that because of careful adherence to ethical considerations that the findings of this study may be applicable and the findings may be consistently repeated with a similar set of criteria (Lincoln & Guba, 1982).

Neutrality

Because I employed a third party to collect data and because I have identified potential biases, the neutrality of the research can be established. The purpose of this study is to understand the impact of feedback and assessment practices on female students in an all-girls high school in an urban, faith-based environment; therefore, the findings will help me better understand the research problem and are not meant to confirm personal perspectives, interests, biases, or motivations (Lincoln & Guba, 1982). According to Creswell and Poth (2018), researchers are ethically bound to “position themselves” in a research study by conveying their background, how it

informs their analysis of the research, and what they might gain from the study (p. 44).

I have acknowledged that as an administrator at the site in which the research was conducted, there is no personal gain from the study; however, the desired professional gain will be to determine future inservice topics for faculty on effective and equitable assessment practices and how female students experience academic feedback and assessment.

Summary

This study used a mixed-methods single case study to determine how female students experience feedback and assessment in an all-girls, faith-based, urban school. This chapter described the purpose of the study, the rationale for a mixed-methods case study, the three-part participant selection criteria, and the data collection and analysis procedures. This explanatory sequential design was conducted in two phases: In Phase 1, quantitative data were collected from teachers and students, including demographic information and responses to survey questions. In Phase 2, qualitative data were collected using a semi-structured interview protocol. Participants were selected for this phase utilizing intensity sampling and interviews were transcribed, coded, and analyzed for findings that expanded on the quantitative data in Phase 1.

Chapter 4: Data & Analysis

The purpose of this case study was to understand the impact of feedback and assessment practices on female students in an all-girls high school in an urban, faith-based environment. This chapter will present the findings of the research organized by the following guiding research questions:

1. What do teachers at an all-girls, faith-based urban high school perceive to be formative and summative assessment strategies that support improved student learning in their classroom?
2. How do female students experience feedback and assessment practices in an all-girls, faith-based, urban high school?

Research Question #1: Teachers' Self-Perceived Skills with Feedback and Assessment

To address the first research question, understanding teachers' self-perceived skill on assessment strategies that support improved student learning, 43 faculty members completed the Scale of Teacher Assessment Practices (STAP) survey (Howell, 2013), including 11 men (26%) and 32 women (74%), 34 of whom were white (79%), five were people of color (11%), and four preferred not to specify (9%). The level of experience spanned a wide range, with 33% of faculty teaching between 0-9 years, 30% teaching between 10-20 years, and 37% teaching 21 or more years. Almost half (49%) of the participants reported never having taken a course in assessment, and 40% of the participants reported never having attended an inservice specifically focused on training in assessment. Demographic information is summarized in Table 4 including teacher race/ethnicity.

Table 4*Demographic Characteristics of Participants (n = 43)*

Characteristic	<i>n</i>	%
Gender		
Male	11	26
Female	32	74
Number of Years Teaching		
0-4	11	26
5-9	3	7
10-14	6	14
15-20	7	16
21+	16	37
Race/Ethnicity		
Asian/Pacific Islander	1	2
Hispanic/Latino/Latinx	1	2
Multi-Racial	3	7
White	34	79
Prefer Not to Answer	4	9

Scoring the STAP Survey

The six assessment domains on the STAP survey (see Table 3) contained 36 items that were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. Participants could choose “n/a” if they had no knowledge of or experience with an assessment practice. In this analysis, “n/a” was scored “0” with the rationale that there were teachers who participated in the survey who were new to the profession, were newly hired, and had only been teaching a few weeks. These teachers might have no knowledge or experience with a particular assessment practice. Some teacher participants with teaching experience might have only taught in this one school, and since this particular school did not use standardized

testing as an assessment practice, they might have had no experience with using and/or analyzing standardized assessments. Therefore, choosing the “very low” option was not accurate, and the “n/a” option, which carried a score of “0,” was a more accurate score. The highest number a participant could score was 180 and the lowest was 0 (a score of 0 was possible only if a participant chose “n/a” for every item). The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high.” The compression of scores provided the level of detail necessary for this study.

Relationship between Teachers’ Experience and STAP Score

Descriptive statistics were used to determine whether there was a relationship between a teacher’s years of experience and how they rated their assessment skills. As shown in Table 5, the average STAP score for all teacher participants ($n = 43$) was 129.60 ($SD = 20.63$) with a median score of 133. As might be expected, teachers who identified as teaching between 0-4 years rated their skills with assessment practices lower than the other groups. There was little difference in the total STAP score for faculty members who had been teaching five or more years. ANOVA results indicated that none of the groups differed significantly ($p > .05$).

Table 5*Relationship between Years Teaching/Total STAP Score*

Years Teaching	<i>n</i>	<i>M</i> (<i>SD</i>)	Median
0-15+	43	129.60 (20.63)	133
0-4	11	118.82 (19.41)	126
5-14	9	133.00 (20.04)	134
15+	23	133.44 (20.43)	139

Selecting and Developing Assessment Methods

Table 6 focuses on the Selection and Development of Assessment Methods domain. Items in this part of the survey asked teachers to rate their ability to select and develop various assessment methods to meet their assessment goals and their students' needs. Seven items were within this domain, which are included in Table 6.

Table 6*Teachers' Self-Perception of Their Knowledge, Experience, and Skills with Selecting and Developing Assessment Methods*

Question	<i>M</i> (<i>SD</i>)	Median	% Low	% Accept- able	% High	% n/a
Skill with choosing an assessment method for a specific purpose relating to an individual student	3.67 (0.92)	4	9	35	56	0
Skill with selecting multiple methods of assessment (e.g., tests, discussions, observations, etc.)	4.12 (0.79)	4	5	12	84	0
Skill with creating assessments that accommodate the needs of a variety of students	3.65 (0.92)	4	12	30	58	0
Skill with determining if an assessment is aligned with required standards (e.g., departmental or school goals)	3.88 (0.94)	4	7	21	70	2
Knowledge of which externally produced assessments are current and available	2.74 (1.00)	3	44	37	19	0
Skill with developing assessments with different formats (e.g., multiple-choice, fill-in-blank, short answer)	4.23 (0.66)	4	0	12	81	7
Skill with sampling from the domain defined by learning goals to write assessment items	3.45 (0.99)	3	14	35	44	7

Note. The items were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high” for this analysis. Participants could choose “n/a” if they have no knowledge of or experience with an assessment practice; “n/a” is scored “0” in this analysis.

The data in Table 6 revealed that all 43 participants perceived they had a proficient, if not exemplary, understanding of all but externally produced assessments, and over half felt proficient in externally produced assessments, as well. The assessment areas where teachers rated their skills the highest were selecting multiple methods of assessment (84%) and developing assessments with different formats (81%). The area where teachers felt the least knowledgeable was with externally produced assessments (44%). The two areas with the highest percentage of participants that chose the “n/a” response were “skill with developing assessment with different formats (e.g., multiple-choice, fill-in-blank, short answer)” and “skill with sampling from the domain defined by learning goals to write assessment items.” Participants who chose “n/a” for these two areas had various years of experience; therefore, there was no apparent connection between their responses and their years of service, which may in part be due to the small sample size ($n = 43$). Table 6 underscores that 43 participants perceived that they had a proficient understanding of all but externally produced assessments, yet on a question about teachers’ training in assessment in the demographic portion of the survey, half reported that they have never had an assessment course or any inservice training specifically focused on assessment practices.

Administering, Scoring, and Interpreting Assessment Results

The six items in this part of the survey asked teachers to rate their skills with administering both formative and summative assessments, as well as their ability to score and interpret the results. The results are presented in Table 7.

Table 7

Teachers' Self-Perception of Their Knowledge, Experience, and Skills with Administering, Scoring, and Interpreting Assessment Results

Question	<i>M</i> (<i>SD</i>)	Median	% Low	% Accept- able	% High	% n/a
Skill with seeking assistance when I am unsure how to score an item	4.17 (0.90)	4	5	19	74	2
Skill with interpreting summary scores reported with standardized test results (e.g., mean, percentile rank on tests such as HSPT, AP, etc.)	3.26 (0.99)	3	21	40	30	9
Skill with administering progress-monitoring (formative) assessments	3.84 (0.84)	4	2	37	60	0
Skill with administering standardized assessments (e.g., standardized achievement tests)	3.28 (1.02)	3	12	47	33	9
Skill with interpreting criterion-referenced scores is (criterion-referenced scores make a statement about how well a student performs, regardless of how other students perform)	3.05 (1.06)	3	28	33	28	12
Skill with understanding why standardized administration is necessary to interpret results of standardized tests	3.26 (1.27)	3	26	26	40	9

Note. The items were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high” for this analysis. Participants could choose “n/a” if they have no knowledge of or experience with an assessment practice; “n/a” is scored “0” in this analysis.

The data in Table 7 reveal that teachers at the participating school felt most comfortable with both formative assessment practices and seeking assistance when determining how to score an assessment. Teachers felt the least skilled in administering and interpreting standardized assessments. Teachers perceived themselves to be highly skilled with seeking assistance when they were unsure of how to score an item (74%), as well as their skill with administering progress-monitoring (formative) assessments (60%). On items related to standardized assessment, teachers more often reported that they perceived their skills to be low, which could be because the setting was a private school and other than A.P. classes, only teachers with experience in other school settings might have been familiar with a standardized assessment. For example, 28% of teachers rated their skills low with “interpreting criterion-referenced scores (scores that make a statement about how well a student performs, regardless of how other students perform). In addition, this domain’s highest percentage of teachers (12%) chose “n/a” for this item, which meant they had no knowledge of or experience with interpreting criterion-referenced scores. The five participants who chose “n/a” represented a range of experience levels. Overall, teachers chose “n/a” more often for this domain compared with the other five domains on the survey.

Using Assessment Results to Inform Day-to-Day Decisions

Assessment items in this part of the survey asked teachers to rate their skills with using the results of both formative and summative assessments to make decisions about lesson content and delivery, to monitor learning, to adapt lessons based on

assessment results, and to group students based on learning needs. Six items were within this domain, which are included in Table 8.

Table 8

Teachers' Self-Perception of Their Knowledge, Experience, and Skills Using Assessment Results to Inform Day-to-Day Decisions

Question	<i>M</i> (<i>SD</i>)	Median	% Low	% Accept- able	% High	% n/a
Skill with using the results of formative assessment to adjust the content of my lessons	4.12 (0.73)	4	0	21	79	0
Skill with using assessment results to appropriately group students for instruction	3.74 (0.91)	4	7	30	53	9
Skill with using results of summative assessments to adjust future lesson plans	4.09 (0.72)	4	0	21	79	0
Skill with using assessment information to develop an instructional plan for a student	3.65 (0.61)	4	2	35	63	0
Skill with using progress monitoring results (formative assessment) to adjust instruction	3.79 (0.78)	4	2	35	60	2
Skill with using assessment results to identify students with similar learning needs	3.60 (0.86)	4	12	28	58	2

Note. The items were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high” for this analysis. Participants could choose “n/a” if they have no knowledge of or experience with an assessment practice; “n/a” is scored “0” in this analysis.

Table 8 shows that all participants perceived their skills to either be acceptable (21%) or high (79%) in both skill in using the results of formative assessment to adjust the content of their lessons and skill in using results of summative assessments to adjust future lesson plans. This domain had few participants who rated their skills low or indicated they had no knowledge of or experience (“n/a”) in the assessment areas. The data in Table 8 show that five participants (12%) did not consider themselves proficient with using assessment results to identify students with similar needs. When looking closer at these five participants’ demographic information, only one of the teachers was a new teacher (0-4 years of experience). The other participants had over 10 years of teaching experience. Identifying students with similar needs is an important component in teaching and learning, and the data revealed an opportunity for future professional development in this assessment area.

Communicating Assessment Results to Others

Items in this part of the survey asked teachers to rate their ability to effectively communicate assessment results to students, parents, and fellow educators. Six assessment items were included within this domain and are set forth in Table 9.

Table 9

Teachers' Self-Perception of Their Assessment Practices Knowledge, Experience, and Skills with Communicating Assessment Results to Others

Question	<i>M</i> (<i>SD</i>)	Median	% Low	% Accept- able	% High	% n/a
Skill with explaining assessment results clearly to parents	3.88 (0.80)	4	5	23	70	2
Skill with selecting appropriate methods for reporting results to others in addition to grades	3.46 (0.72)	3	5	47	40	9
Skill with explaining results to other educators for the purpose of assisting with placement decisions	3.95 (0.81)	4	5	19	70	7
Skill with explaining to parents how assessment results are used to make decisions about their children	3.52 (0.91)	3.5	12	35	47	7
Skill with communicating the results of assessments to students in a way that they can understand	3.95 (0.82)	4	2	28	70	0
Skill with explaining to students how assessment results will be used to assign grades	4.00 (0.76)	4	2	21	77	0

Note. The items were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high” for this analysis. Participants could choose “n/a” if they have no knowledge of or experience with an assessment practice; “n/a” is scored “0” in this analysis.

The data in Table 9 show that in four of the six assessment items, 70% or more of the teachers considered themselves highly skilled. The lowest scoring item in this

domain was teachers' perceived skill in explaining to parents how assessment results are used to make decisions about their children. It is unclear from the data if the participants who rated their skills low on this assessment item did so because they did not feel skilled in communicating the results to parents or that they lacked proficiency in using assessment results to make decisions. Regardless, communicating assessment results to parents proved difficult. Notably, the three participants who selected "n/a" for this item all had 10 or more years of teaching experience.

Ethical Use of Assessment

Items in this part of the survey asked teachers to rate their skill in using assessment appropriately, communicating assessment results in an ethical manner, and understanding their legal responsibility concerning assessment. Five assessment items were included within this domain, as set forth in Table 10.

Table 10

Teachers' Self-Perception of Their Assessment Practices Knowledge, Experience, and Skills with Ethical Use of Assessment

Question	<i>M</i> (<i>SD</i>)	Median	% Low	% Accept- able	% High	% n/a
Skill with adhering to the bounds of confidentiality regarding assessment results	4.60 (0.66)	5	0	9	91	0
Knowledge of the consequences of unethical use of assessment	3.51 (1.07)	4	23	23	53	0
Skill with recognizing inappropriate use of assessment	3.74 (0.89)	4	9	26	63	2
Experience with recognizing when assessment results are being used inappropriately by others	3.08 (0.99)	3	26	37	26	12
Knowledge of identifying my own legal responsibilities in regard to assessment	3.16 (1.09)	3	23	37	40	0

Note. The items were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high” for this analysis. Participants could choose “n/a” if they have no knowledge of or experience with an assessment practice; “n/a” is scored “0” in this analysis.

The items in this area of assessment focus on the ethical use of assessment and teachers' responsibility to use assessment data appropriately. Table 7 shows that teachers perceived themselves to be highly skilled with adhering to the bounds of confidentiality regarding assessment results (91%), with no participants viewing their skills as low. The next highest-ranking item in this domain is skill with recognizing inappropriate use of assessment, with 63% of participants rating their skills as high.

The assessment item in this domain where teachers felt the least knowledgeable was experience recognizing when assessment results were being used inappropriately by others (26%). Five participants chose “n/a” for this assessment item, which was the highest percentage of “n/a” scores in the domain.

This domain had the highest percentage of participants rating their skills “low” (16%), which raises the question about whether participants did not have experience with ethical use of assessment or whether they had experience and still did not believe they were proficient. The data in Table 7 reveal possible professional development opportunities specifically focused on appropriate uses of assessment data.

Overall Skill with Assessment Practices

The assessment items in this part of the survey focused on teachers’ self-perception of their overall skills with assessment practices, which encompassed a broader look at assessment compared with the other domains. Six overall assessment items were within this section, which are included in Table 11.

Table 11*Teachers' Self-Perception of Their Overall Skill with Assessment Practices*

Question	<i>M</i> (<i>SD</i>)	Median	% Low	% Accept- able	% High	% n/a
Skill with writing fill-in-the-blank/short answer questions	3.95 (0.74)	4	2	21	72	5
Skill with using assessment results when developing lesson plans	3.88 (0.76)	4	5	21	74	0
Skill with revising a test based on item analysis	3.59 (0.86)	4	7	28	51	14
Skill with using assessments such as classwork to enhance my instructional delivery	4.33 (0.68)	4	0	12	88	0
Skill with using assessment results to improve teaching and learning	4.05 (0.82)	4	5	16	79	0
Skill with developing assessments based on clearly defined course objectives	3.93 (0.77)	4	2	26	72	0

Note. The items were scored using a five-point Likert-type scale with response choices ranging from “1” very low to “5” very high. The “1” and “2” scores were combined and scored “low” and the “4” and “5” scores were combined and scored “high” for this analysis. Participants could choose “n/a” if they have no knowledge of or experience with an assessment practice; “n/a” is scored “0” in this analysis.

The items in this part of the survey focused on teachers’ overall assessment skill. The data in Table 11 show that in five of the six assessment items, 72% or more of the participants considered themselves highly skilled. Teachers self-perceived skill with revising a test based on item analysis was the only item scoring less than 72%, with 51% of participants considering themselves highly skilled. This item had the

most participants selecting “n/a” (14%), which suggests that this is the area where the highest number of teachers had no experience with or knowledge of that assessment practice. All participants considered their skills “acceptable” or “high” in using assessments like classwork to enhance instructional delivery.

Summary of Research Question #1

To understand the impact of feedback and assessment practices on female students in an all-girls high school, it is important to learn how teachers perceive their knowledge, experience, and skills in using assessments to evaluate student work. Participants completed a survey that included questions about their assessment practices, demographic information, and assessment training and coursework, as well as whether they included non-academic elements in summative grades.

A repeated measures ANOVA was run to compare the means between the six assessment domains. Table 12 displays the means and standard deviations for each of the six domains in the survey, and the analysis of variance showed that there was a statistically significant difference by domain ($p < .05$). Post hoc tests revealed that teachers rated themselves significantly lower on Administering, Scoring, and Interpreting Assessment Results and rated themselves the highest on Overall Assessment Skills.

Table 12*STAP Items by Domain*

STAP Domains	STAP Items	<i>M</i> (<i>SD</i>)
Selection and development of assessment methods	3, 11, 15, 17, 18, 27, 29	3.59 (.62)
Administering, scoring, and interpreting assessment results	2, 13, 14, 19, 25, 26	3.25 (.83)
Using assessment results to inform day-to-day decisions	4, 6, 9, 22, 23, 24	3.74 (.63)
Communicating assessment results to others	1, 7, 8, 16, 21, 30	3.64 (.75)
Ethical use of assessment	5, 10, 12, 20, 28	3.53 (.75)
Overall skill with assessment	31, 32, 33, 34, 35, 36	3.84 (.65)

The data from the portion of the survey that included the Scale of Teacher Assessment Practices survey show that teachers reported being highly skilled with writing assessments, using assessment results to guide instruction and improve learning, and communicating assessment results. By comparison, many faculty members on the demographic portion of the survey reported not having current training in assessment practices, with 40% reporting having never attended an inservice on assessment and 49% reporting having never taken a class on assessment. A large percent of the faculty (60%) reported including non-academic elements in summative grades, and the majority of faculty members (95%) expressed interest in training specifically focused on equitable assessment.

Research Question #2: Female Students' Experience with Academic Feedback

To address the second research question about how female students experience feedback and assessment practices in an all-girls, faith-based, urban high school, student volunteers were recruited from the 11th and 12th grades to participate in a survey, the Instructional Feedback Orientation Scale (IFOS) (King et al., 2009). This survey was designed to measure students' perceptions of how they experience academic feedback. There was no incentive other than to participate in a survey. There were 101 participants (56 juniors, 45 seniors), and 22% of participants identified as students of color and 78% identified as White. The 27-item student survey asked participants to rate how they experience grades and academic feedback in four dimensions, namely utility, sensitivity, confidentiality, and retention. Each response was on a five-point Likert-type scale where "strongly disagree" was "1" and "strongly agree" was "5." The "1" and "2" scores were combined and scored "disagree," the "3" score is neutral, and the "4" and "5" scores were combined and scored "agree" for this analysis. Because "3" represents "neither agree nor disagree," this option might be a limitation of the student results because it does not require participants to indicate their level of agreement with a statement.

IFOS: Utility of Grades and Academic Feedback

Table 13 focuses on the Utility dimension of the IFOS, which contains 10 questions about the extent to which students found feedback useful.

Table 13*Students' Perception of Feedback Utility*

Question	<i>M</i> (<i>SD</i>)	Median	% Disagree	% Neutral	% Agree
I think feedback from teachers is vitally important in improving my performance.	4.45 (0.52)	4	0	1	99
I will usually reflect on a teacher's feedback.	4.17 (0.58)	4	1	7	92
I listen carefully when a teacher provides feedback.	4.16 (0.67)	4	1	13	86
I am extremely encouraged by positive feedback from teachers.	4.56 (0.68)	5	1	8	91
I think that feedback provides clear direction on how to improve my performance.	3.97 (0.82)	4	5	20	75
Feedback from my teachers can be a valuable form of praise.	4.18 (0.75)	4	2	15	83
I pay careful attention to instructional feedback.	4.09 (0.62)	4	1	12	87
Feedback from my teachers motivates me to improve my performance.	3.89 (0.93)	4	8	25	67
Feedback from teachers is a waste of time.*	1.59 (0.65)	2	91	9	0
I feel relieved when I receive positive feedback.	4.58 (0.53)	0	0	2	98

Note. *Item is reverse coded; each response was on a five-point Likert-type scale where "strongly disagree" was "1" and "strongly agree" was "5." The "1" and "2" scores were combined and scored "disagree," the "3" score is neutral, and the "4" and "5" scores were combined and scored "agree" for this analysis.

Student participants overwhelmingly agreed (99%) that feedback from teachers was vitally important in improving their performance. Similarly, students noted that they were extremely encouraged by positive feedback (91%), and 98% reported feeling relieved when they received positive feedback. However, the percentage drops greatly (67%) when students were asked if feedback from teachers motivates them to improve (67%) and if feedback provides clear direction on how to improve (75%). These data are important to note given that 88% of teachers considered themselves highly skilled in using assessments like classwork to enhance their instructional delivery and 79% considered themselves highly skilled in using assessment results to improve teaching and learning (see Table 11).

IFOS: Sensitivity to Grades and Academic Feedback

Table 14 focuses on the Sensitivity dimension of the IFOS, which contains nine questions about the extent to which students were sensitive to teacher feedback.

Table 14*Importance of Considering Students' Sensitivity to Feedback*

Question	<i>M</i> (<i>SD</i>)	Median	%Disagree	%Neutral	%Agree
My feelings can be easily hurt by corrective feedback from a teacher.	3.07 (1.06)	3	35	30	36
I feel threatened by corrective feedback.	2.36 (0.99)	2	62	23	15
Corrective feedback hurts my feelings.	2.45 (0.94)	2	58	26	16
Corrective feedback is intimidating.	3.27 (1.06)	3	29	24	48
My feelings are not easily hurt by corrective feedback from a teacher.*	3.28 (0.95)	3	24	27	50
It is difficult to “get over” corrective feedback.	2.75 (1.06)	3	49	21	31
Corrective feedback is embarrassing.	2.92 (1.11)	3	36	31	34
I tend to dwell on the negative feelings that result from corrective feedback.	3.33 (1.03)	3	24	29	48
Corrective feedback from a teacher increases the stress I feel about future performance.	3.47 (1.11)	4	24	17	59

Note. *Item is reverse coded; each response was on a five-point Likert-type scale where “strongly disagree” was “1” and “strongly agree” was “5.” The “1” and “2” scores were combined and scored “disagree,” the “3” score is neutral, and the “4” and “5” scores were combined and scored “agree” for this analysis.

While relatively few participants agreed that they felt threatened by corrective feedback (15%) or that corrective feedback hurt their feelings (16%), 59% of participants noted that corrective feedback from a teacher increased the stress they felt about future performance. This question yielded the highest scoring in the category, followed by questions noting that corrective feedback was intimidating (48% agreed) and that participants tended to dwell on the negative feelings that resulted from corrective feedback (48% agreed).

The treatment of the “neutral” option for the items in Table 14 may require special consideration, as items marked “neutral” related to sensitivity to teacher feedback may indicate a stronger alignment with the “agree” option given the nature of the questions. Depending on how “neutral” is interpreted, 66% of participants indicated that they feel they feel hurt or may feel hurt by a teacher’s corrective feedback. In addition, depending on how the “neutral” option is interpreted, 65% of participants indicated that corrective feedback is embarrassing. This is important to consider given the data presented in Table 8 indicating teachers consider themselves highly skilled in giving feedback.

IFOS: Importance of Confidentiality with Grades and Academic Feedback

Table 15 focuses on the Confidentiality dimension of the IFOS, which contains five questions related to how student participants preferred to receive academic feedback.

Table 15*Importance of Confidentiality in Receiving Academic Feedback*

Question	M (SD)	Median	%Disagree	%Neutral	%Agree
I do not like to receive corrective feedback in front of other people.	4.16 (0.87)	4	5	16	79
I do not like for others to hear what feedback I am receiving.	4.14 (0.81)	4	5	12	83
I don't mind being singled out by feedback from a teacher.*	2.31 (1.03)	2	62	22	16
I prefer to receive feedback from a teacher in private.	4.25 (0.71)	4	2	10	88
I like others to hear the feedback I am receiving from my teacher.*	1.83 (0.76)	2	82	16	2

Note. *Item is reverse coded; each response was on a five-point Likert-type scale where “strongly disagree” was “1” and “strongly agree” was “5.” The “1” and “2” scores were combined and scored “disagree,” the “3” score is neutral, and the “4” and “5” scores were combined and scored “agree” for this analysis.

The highest scoring question in this dimension reveals that 88% of participants preferred to receive feedback from a teacher in private. The description of this question does not define “in private,” so it is unclear whether “in private” means in person and in private, via email in private, or just not publicly announced in some manner. The three other questions related to receiving feedback privately had similar scores ranging from 79%-83%. The question asking participants if they minded being singled out by feedback had the lowest agreement with 62% noting that they do not mind being singled out. The data in the two items that were reverse coded aligned with

the participants' responses to the same items that were expressed in the inverse. For example, 2% of participants disagreed that they preferred to receive feedback from a teacher in private and 2% agreed that they liked others to hear the feedback they received from their teacher. One limitation of the data presented in Table 15 is that "feedback" may be interpreted as only negative feedback. It would be interesting to note whether students would be more receptive to teachers giving student feedback in front of others.

IFOS: Retention of Grades and Academic Feedback

Table 16 focuses on the Retention dimension of the IFOS, which contains three questions (each stated in the negative) about whether student participants cared about teacher feedback and how they comprehended and retained feedback.

Table 16*Students' Retention of Academic Feedback*

Question	M (SD)	Median	%Disagree	%Neutral	%Agree
I can't remember what teachers want me to do when they provide feedback.	2.77 (0.94)	3	46	29	26
I tend to miss out on the details of what instructors want when they provide me with feedback.	2.84 (0.99)	3	45	26	30
I typically do not make note of the teacher's corrective comments.	2.24 (1.00)	2	71	14	15

Note. Each response was on a five-point Likert-type scale where "strongly disagree" was "1" and "strongly agree" was "5." The "1" and "2" scores were combined and scored "disagree," the "3" score is neutral, and the "4" and "5" scores were combined and scored "agree" for this analysis.

The highest scoring question in the Retention dimension of the IFOS focused on whether participants took the time to consider feedback, and 71% reported that they do make note of teacher's corrective comments. More than half of the participants reported either being "neutral" or agreeing with being unable to remember what teachers want them to do when they provided feedback (55%) and missing out on the details of what instructors want when they provided feedback (56%). These data illuminate important considerations for teachers. Just over 70% of participants reported that they make note of teachers' feedback. Yet just over half (55%) reported that they do not remember what teachers want them to do differently and that they miss the important details of the feedback (56%). One potential barrier to interpreting these findings could be that these questions were phrased with a negative sentence

construction (“I can’t remember,” “I tend to miss out,” “I typically do not make note”), and participants might have misunderstood the intent of the questions.

However, if students accurately interpreted the questions, it would mean that one-half of the participants make note of teachers’ feedback, yet they do not understand or remember teachers’ recommendations about how they could improve.

Student Interviews

To learn more about how female students in an urban, faith-based high school experience grades and academic feedback, purposive sampling was used to identify students willing to participate in semi-structured interviews. The juniors and seniors who participated in the survey were sent an email introducing the survey and requesting they respond to the email if they were interested in being selected for a follow-up interview. Of the group who participated in the IFOS survey, 10 students (two juniors and eight seniors) responded to the email, and all 10 students were selected to participate in the interviews. Student participants were asked to bring a graded essay, project, or exam with them to discuss. During the interview, they were asked to describe the assignment, why they chose it, the feedback they received, and how they experienced the feedback. Finally, students described to what extent, if any, the feedback impacted their future performance, academic choices, and academic self-efficacy, the degree to which one perceives their ability to learn or perform behaviors at a designated level (Bandura, 1997; Schunk & Pajares, 2005). A neutral third party conducted the interviews, which were recorded on Zoom. Each interview lasted approximately 13-15 minutes.

Findings. The interviews were evaluated to learn students' perspectives, in their own voices, about how they experienced academic feedback and grades. During the first-cycle coding process, the transcripts from each interview were reviewed numerous times and were coded using the holistic, emotion, and In Vivo coding approaches (Saldaña, 2018). After analyzing the codes from the first cycle, some outliers were eliminated because they only appeared in a single interview. Some codes that appeared infrequently were combined during the second-cycle coding process. The remaining 104 holistic, emotion, and In Vivo codes were analyzed in cycle two (pattern coding) resulting in eight categories, which were further analyzed into the following three themes: Assignment Type, Feedback/Grading, and Academic Self-Efficacy/Self-Esteem, as shown in Table 17. According to Bandura (1997), self-efficacy deals with students' perception of their ability to learn, while self-esteem centers around someone's sense of value and self-worth (Villeneuve et al., 2019). For the purposes of this study, these two categories were combined as they focus on students' perception of who they are, what they can do, and how they feel. While the themes are introduced in this chapter, how the data and associated findings relate to these themes will be discussed in more depth in Chapter 5.

Table 17*Themes and Categories in Second-Cycle Coding Analysis*

Theme	Categories
Assignment Type	Summative Assessment Formative Assessment
Feedback/Grading	Positive Feedback Neutral Feedback Negative Feedback Grades/Grading
Academic Self-Efficacy/Self-Esteem	Academic Self-Efficacy Self Esteem

Theme: Assignment Type. Students were given instructions to bring one or two graded assignments to talk about during the interview, such as a project, test, essay, lab, or any longer assignment. Instructions also noted that work such as math homework or vocabulary quizzes would not be good choices to discuss for the interview because these assignments might not contain enough detailed feedback. Table 18 describes the reasons participants chose their assignments in their own words.

Table 18

Responses to Question: "Tell Me Why You Chose This Assignment to Talk About"

Positive/ Negative	In Vivo Code
Positive	<p>I put a good amount of effort into this assignment.</p> <p>Made me feel really good that it wasn't just focusing on the negative.</p> <p>I'm really proud of the work that I did.</p> <p>I can look at this and say, OK, here's what I need to do better because I can build on this.</p> <p>It didn't just say "good work." It said, "I like that you did this." I like that she gave a specific example of what she likes.</p> <p>This assignment is one that I've gotten the most feedback on.</p> <p>She wrote really good comments on it of what I needed to work on.</p> <p>I feel like this was one of those essays that was kind of my best work, but also the work I got the most feedback from.</p>
Negative	<p>It wasn't the most satisfying feedback I could have received because it was also very general. It wasn't specific about what I need to do. . .it's kind of vague and it was harder to understand what I needed to fix.</p> <p>I chose it mostly because it's the first time I've ever gotten a "C" since 7th grade math. And also, when I first saw this grade, I immediately had to go to the counselor's office because I wanted to kill myself.</p>

The data in Table 18 reveal that participants were interested in discussing assignments from the Humanities, including a variety of essays for English and a video project for American Government. It is interesting that while the parameters

allowed for students to bring in graded work from any of the disciplines, they chose similar assignments to discuss. This could be due to a number of reasons, such as teachers in the Humanities generally offer more feedback because assignments are more subjective in nature and require more explanation for improvement, or that students feel more connected to their writing and therefore the feedback feels more personal to them.

The data in Table 18 show that eight of the 10 participants chose to discuss assignments that had positive feedback or feedback that impacted their emotions in a positive manner. Two of the 10 student participants chose to discuss assignments where they felt they received vague or inadequate feedback or feedback that negatively impacted their emotional well-being. Researchers who choose to repeat this study might consider amending the instructions so that students bring one assignment containing positive feedback and one containing negative feedback. An interview where students have an opportunity to discuss both positive and negative feedback might result in more complete findings on students' experiences with feedback.

Theme: Feedback/Grading. The questions on academic feedback elicited more detailed responses from the participants compared with the other interview questions. There were 21 holistic codes, 9 emotion codes, and 32 In Vivo codes that emerged from the data. The interview transcripts were then reviewed with a specific focus on questions pertaining to students' experiences with the feedback and/or grades, including: "Was there feedback on this that you found especially helpful or encouraging?" and "Was there feedback that you found discouraging or unfair/unjust?" The resulting In Vivo codes were categorized as positive, neutral, or

negative based on how the student experienced the feedback. For example, the comment “I felt like my teacher provided really good feedback that I can look over and it's very clear and helpful” was grouped in the Positive Feedback category. The In Vivo codes that had to do with future improvement like, “I can look at this and say, okay, here's what I need to do better because I can build on this,” were grouped in the Neutral Feedback category. Comments on feedback like, “teachers who don't grade for months at a time and then we have five essays to complete in that time . . . I'm not really learning from the feedback because I don't get the feedback until the end of the semester” were In Vivo code classified as Negative Feedback. The literature suggests that the least effective forms of feedback were overly general comments, whether or not the comments consisted of praise or punishment (Hattie & Timperley, 2007; Harks et al., 2014), and the comments from the student participants support this, as shown in the “Negative” category in Table 19.

Table 19*Responses to Questions about Academic Feedback.*

Feedback	In Vivo Code
Positive	<p>It blew me away. It's some of the best feedback I've ever gotten.</p> <p>One thing was how long it [the feedback] was.</p> <p>I feel like whenever a teacher shows that they care, it makes all the students want to care more.</p> <p>She graded assignments very quickly and got them in, got them back to you within a few weeks, which was always super helpful.</p> <p>Feedback can be the best parts of an education and it's what really helps with growth.</p>
Neutral	<p>Whereas hearing what things, specifically, he wanted to hear about more, I know to focus more on that next time</p> <p>It's really satisfying to not see a big red X; the absence of negative feedback feels nice.</p> <p>Here's how you can improve in the future. . .here's what to look for next time. . .here's what to do to get there.</p>
Negative	<p>I don't know what I got wrong or what to work on.</p> <p>I'm not really learning from the feedback because I don't get the feedback until the end of the semester.</p> <p>If it's bad, I'm normally more likely to take feedback harder because I'm scared and if it's not perfect.</p>

The interview responses specific to the final grade on the assignment pertained to whether or not the teacher provided a rubric in advance supported the literature, which underscores the benefits of transparent grading practices (Feldman, 2019). The data also revealed students' dissatisfaction when they were confused as to why points were deducted from their scores. Some examples of In Vivo codes specific to grades

include: “Teacher gave me a rubric to look at, and so that was really helpful,” “There wasn't a benchmark,” and “Why are you taking them off if you can't tell me what I'm missing here?”

Theme: Academic Self-Efficacy/Self-Esteem. The literature underscores the connection between grades and academic self-efficacy and self-esteem (Villeneuve et al., 2019; Twenge et al., 2019), and five of the participants (50%) indicated a relationship between academic feedback and grades to their self-efficacy and/or self-esteem during their interviews. This is consistent with the data collected on the Sensitivity dimension of the Instructional Feedback Orientation Scale survey, which notes that 48% of the participants agreed or strongly agreed that corrective feedback is intimidating, and 48% noted that they agree or strongly agree that they dwell on the negative feelings that result from corrective feedback. More than one-third of participants (36%) agreed or strongly agreed that their feelings can be easily hurt by corrective feedback, and 34% noted that corrective feedback is embarrassing.

Participants' responses to the interview question asking them to explain the reasons they chose the assignment resulted in many valuable data points also related to the self-esteem/self-efficacy category. As shown in Tables 18 and 19, a number of students chose to discuss assignments based on their emotional attachment to the assignment and/or the feedback, or the impact the feedback had on their self-efficacy. They commented that the chosen assignments “made me feel really good,” “I'm really proud of,” “[are] kind of my best work,” and “the absence of negative feedback feels nice,” and “I've gotten some other assignments back with low grades. . .it just knocked my confidence.” One interview participant indicated a dangerous relationship between

her grades and her self-esteem, noting that “when I first saw the grade, I immediately had to go to the counselor’s office because I wanted to kill myself.” While this response is extreme, it is consistent with the literature that indicates the strong connection between grades, anxiety, and mental health in general (Feldman, 2019; Mohanty & Jena, 2015; Villeneuve et al., 2019). The literature reveals that the pressure on high-school students to achieve, even if the pressure is self-inflicted, may come at the expense of physical and emotional health (Feldman, 2019; Villeneuve et al., 2019).

Summary of Research Question #2

To understand the impact of feedback and assessment practices on female students in an all-girls high school, it is important to find out how female students experience these practices. Participating students completed the Instructional Feedback Orientation Scale survey, an instrument designed to measure how they experience instructional feedback in terms of usefulness of feedback, sensitivity to feedback, the manner in which they receive feedback, and their retention of feedback. In addition to the survey, students participated in semi-structured interviews, which provided additional qualitative data to better understand how female students experience feedback and assessment practices.

Summary

This chapter discussed findings derived from data analysis from two participant groups, faculty and students, using three instruments, the Scale of Teacher Assessment Practices survey, the Instructional Feedback Orientation Scale survey, and semi-structured student interviews. Data were organized by research question with

more emphasis on feedback from students. The questions in the student interviews overlapped with two of the dimensions in the student surveys, specifically questions regarding the Utility of Feedback and Sensitivity to Feedback. Findings from the student interviews supported the data collected in the overlapping survey dimensions indicating that students prefer timely and specific feedback, regardless of whether they perceive it as positive or negative; students prefer feedback in addition to a rubric; and students view feedback as a critical component in their future academic improvement. The next chapter will discuss the relationship between how students process academic feedback and potential opportunities for teachers to adapt instructional practices based on the findings. Results will be discussed in relation to existing literature, especially as they relate to equitable grading practices, the impact of academic feedback and grades on self-esteem, and the ways in which female students process academic feedback.

Chapter 5: Discussion

The purpose of this case study was to understand the impact of feedback and assessment practices on female students in an all-girls high school in an urban, faith-based environment. This chapter will present a summary discussion of the findings, draw connections to the research literature, identify limitations of the study, make recommendations for educational practice, and outline implications for future research.

To better understand how female students experience academic feedback and assessment, the perceptions examined included both faculty and students in one faith-based, all-girls, urban high school in the Pacific Northwest. The study included 43 faculty members who were surveyed about their perceived skill level in assessment specifically in the areas of: selection and development of assessment methods; administering, scoring, and interpreting assessment results; using assessment results to inform day-to-day decisions; communicating assessment results to others; ethical use of assessment; and overall skill with assessment. The study also gathered data from 101 11th- and 12th-grade students and investigated their perception of academic feedback and grades, specifically in the areas of the usefulness of feedback, their sensitivity to feedback, their perceptions about confidentiality in how they receive feedback, and their retention of specific feedback. In addition to student survey data, 10 students participated in semi-structured interviews where they were asked questions about how they experienced feedback on a specific assignment of their choice, as well as their experiences with feedback in general. This served to triangulate the data with survey responses and the literature. The findings from faculty

and student responses, considered together in relation to assessment literature, give a richer understanding of how female students experience academic feedback and assessment within this context and provide fodder for future research into the impact of teacher assessment practices more broadly.

Discussion of Findings

For the purposes of this discussion, examining the aggregated findings from both surveys and the student interviews provides a more complete picture of teachers' perceptions and students' perceptions. Analyzing the findings in this manner allowed for three distinct "lived experiences" to emerge from the intended and perceived impact of assessment techniques, which are:

1. teacher training in assessment vs. their perception of assessment skill,
2. giving academic feedback vs. experiencing academic feedback: Intent vs. impact, and
3. considering students' emotional connection to their work in overall feedback and assessment strategies

The findings are organized by these experiences below.

Teacher Training in Assessment vs. Their Perception of Assessment Skill

Results from the data indicate that there is a gap between teachers' perceived skill level with assessment and their participation in preservice and/or inservice training directly focused on assessment practices. Almost half of the participating teachers reported having never taken a course in assessment, and 40% reported never having attended an inservice focused on assessment practices. Yet, surprisingly, more than 72% of the teachers in this study considered themselves skilled or highly skilled

in their overall assessment abilities. As reported in the literature review, current research underscores that teachers experience stress, confusion, and job dissatisfaction with the demands of assessment and providing effective feedback, as well as their level of training and support specifically focused on feedback and assessment practices (Feldman, 2019). A study conducted by Battistone et al. (2019) underscores that a preponderance of the teachers surveyed reported that their only formal training in assessment occurred in their preservice programs before they began their official teaching careers, and the teachers' expressed concerns about the reliability and validity of their grading practices. Despite having little to no formal training in assessment, teachers still perceive themselves to be skilled or highly skilled in many traditional and current practices.

The gap between training in assessment practices and perceived skill level is not confined to teachers new to the field. Five of the seven participating teachers who indicated on the Scale of Teacher Assessment Practices (STAP) survey (Howell, 2013) that they have low skills in "using assessment results to identify students with similar learning needs" were teachers who had 10 or more years of experience and two of the five had 21 or more years of experience. These data could be troubling, especially when 12% of the participants indicated they have low skills with "explaining to parents how assessment results are used to make decisions about their children." These data could indicate that almost one-fifth of participants may not be using assessment results to guide their instruction, or that one-fifth of participants do not understand assessment data well enough to confidently communicate the results of student assessment to parents. Further, if study data are interpreted to mean that 12%

have low skills in using assessment results to guide instruction, it might indicate that five of the 43 teachers continue to assign work and assess in ways that do not take into account how well the students have previously performed, if students are improving or if students possess the prerequisite skills necessary to complete the task accurately. Additional training may help improve practice so that teachers understand and are confident that their feedback and assessment strategies are improving student learning.

Giving Academic Feedback vs. Experiencing Academic Feedback: Intent vs. Impact

While teachers in this setting might not consider the disparity between their training in assessment practices and their perceived skill level problematic, based on the data collected for this work, it appears likely that they might not realize how students experience their feedback. Teachers reported providing academic feedback believing that the feedback provided would help students improve their skills and learning; however, over 30% of the students surveyed reported they find it difficult to “get over” corrective feedback and almost half of the students (48%) reported dwelling on negative feelings. In short, even well-intentioned comments from teachers meant to improve student learning can have an unintended and sometimes negative impact on students. As noted in the literature review, Ryan and Henderson’s (2018) research indicates that if academic feedback evokes negative emotional responses, students are less likely to use the feedback to improve and more likely to reject or ignore comments altogether. In fact, dissatisfaction with teacher feedback is one of the most problematic aspects of a student’s educational experience, and this dissatisfaction has considerable impact on overall achievement (Hattie & Timperley, 2007). To determine the extent to which feedback is effective, Hattie and Timperley (2007)

compared evidence related to feedback from over 500 meta-analyses representing approximately 20 to 30 million students. This analysis examined a number of characteristics and outcomes of feedback, and one conclusion pointed to the necessity of matching the type of feedback with the type of learner and, because of the difficulty of this task, inefficient learners often are provided feedback that does little to improve learning and may even be harmful to the overall learning process (Hattie & Timperley, 2007). Further, feedback that is positive, specific, and task-oriented were the most impactful, and this impact could be more pronounced for female students (Altermatt & Pomerantz, 2003; Hattie & Timperley, 2007; Villeneuve et al., 2019). These findings in the literature are supported by evidence in this work noting that eight of the 10 student participants chose to discuss assignments that had specific, positive feedback. This positive feedback inspired and motivated them, and one student even noted that “the [positive] feedback can be the best parts of an education, and it’s what really helps with growth.” One student noted that “if it’s bad, I’m normally more likely to take feedback harder because I’m scared.” These types of student responses might put additional pressure on teachers in an area of their jobs that is already fraught with frustration. Yet the framing of feedback given (strength versus deficit-based) appears to be as important as the feedback itself if the desired result is student motivation and improvement.

Considering Students’ Emotional Connection to Their Work in Overall Feedback and Assessment Strategies

The data from student participants strongly suggest that students have an emotional connection to their work, especially their writing. Understanding this

connection when providing feedback on assessments is critical to students feeling supported and encouraged and is key to the feedback leading to improvement. When asked to choose an assignment to discuss for their interview, and having freedom to choose assignments from any discipline, 90% of students chose to discuss feedback on essays. Participating students' assignments represented essays they "were really proud of," that represented "a good amount of effort," and "that was kind of my best work." The fact that positive comments resonated with students to a greater degree may indicate that positive comments are a critical element in improving performance, especially for writing-based assessments. Training teachers on this characteristic of how students experience feedback may allow them to offer feedback where their intended impact aligns with the actual impact.

Conclusion of Findings

There were a number of key findings with regard to Research Question #1: What do teachers at an all-girls, faith-based urban high school perceive to be formative and summative assessment strategies that support improved student learning in their classroom? First, the data reveal that teachers perceive they are highly skilled in writing assessments, using assessment results to guide instruction and improve learning, and communicating assessment results. These perceptions might be accurate, and if accurate, the assessment skills, as indicated by the teachers, were not acquired from specific coursework or inservice training in assessment. However, when analyzing specific questions on the faculty survey related to students' experiences, this might not be the case. Nearly half of the teachers (44%) rated their skills low or acceptable in "choosing an assessment method for a specific purpose relating to an

individual student.” Similarly, 42% of teachers rated their skills low or acceptable in “creating assessments that accommodate the needs of a variety of students.” While 60% of teachers believed themselves highly skilled in “using formative assessment to adjust instruction,” this means that close to 40% feel only adequately prepared or unprepared in this area. As evidenced by the item analysis, even when teachers perceive themselves to be highly skilled with assessment, when asked questions specifically related to students’ experience with their assessment practices, their perceived skill level drops precipitously.

There were several key findings with regard to Research Question #2: How do female students experience feedback and assessment practices in an all-girls, faith-based, urban high school? Based on the results of the student survey, students overwhelmingly think “feedback from teachers is vitally important to improving [their] performance” (99%), and they take spend time “reflect[ing] on a teacher’s feedback” (92%). It is evident that teacher feedback is important to students. But the type of feedback and the way it is offered appears to be as important or, in some cases, more important. Receiving positive feedback from teachers leaves 91% of students feeling “extremely encouraged,” and 98% report “feeling relieved when they receive positive feedback.” Almost half of the participating students (48%) are intimidated by corrective feedback and dwell on the negative feelings that result from corrective feedback. The percentage shoots up when asked if corrective feedback makes them feel embarrassed when it is offered in front of others (88%) and 59% feel stress about their future performance after receiving corrective feedback. One student even noted that “if [feedback] is bad, I’m normally more likely to take [it] harder because I’m

scared.” Because providing feedback is a critical component of student learning, and because of the profound impact of feedback and assessment on students, every effort should be provided to train teachers in best practices.

Limitations of the Study

There are four sets of limitations that impacted or may have impacted the data collected for this study.

Limitations Due to COVID-19 Pandemic

The COVID-19 pandemic limited the data, especially as compared with data in other studies not impacted by the pandemic’s extraordinary effect on teaching and learning. In a typical year, students at the participating school would complete two sets of final exams and experience a wide variety of feedback and assessments. But the COVID-19 pandemic eliminated many of the typical ways students demonstrate learning and skills. From March 2020 to February 2021, students learned and were assessed in an online environment. Assessments were modified significantly during this period. From March 2021 to April 2021, students attended classes either completely online or in a hybrid situation where two days per week they attended school in person, two days they attended classes online, and one day per week they opted into online enrichment activities. From April 3, 2021 to June 2021, students were either in-person four days per week or online four days per week (with one day for enrichment). There were no final exams given from March 2020 to June 2021. Goals for assessment are shifting as students and teachers reacquainted themselves with the pace of the school day and set revised expectations for student learning. In a time when mental health and wellness are equally important, it is difficult to know if

and how teachers may have altered their feedback and assessment styles given the new focus on student mental health over academic performance.

Limitations Due to Role of Researcher

Because I am the principal of the school, I am in a role of authority, so to address the power differential, I employed neutral third parties to collect the data from both the faculty and students. The neutral third-party data collector for the faculty was a research fellow with no connection to the school. This allowed faculty to feel greater comfort providing honest responses without fear of repercussions. The neutral third-party data collector for the students was the librarian, who is a trusted member of the community, but who does not directly teach or assess students. Even with this level of separation, faculty members and students may have perceived a connection between the third-party data collector and the researcher. To help mitigate the effects of potential biases, I used bracketing with continual reminders of any preexisting beliefs and assumptions that may influence the data analysis (Creswell & Poth, 2018). Together with bracketing, I used analytic memos, a self-reflective act of writing notes that clarified the role of the researcher and to minimize subjective influences that erode the collection and analysis of the data (Creswell & Poth, 2018). While I made every attempt to minimize the possible impact of positional power, it is possible that the influence existed.

Limitations Due to “Neutral” Option on Student Survey

One of the limitations of the student survey, the Instructional Feedback Orientation Scale (IFOS) (King et al., 2009), was the “neutral” option, because the way some of the neutral-response data were treated on a number of the questions is

significant in determining how students experienced feedback. The IFOS allowed students five response choices ranging from “1” strongly disagree to “5” strongly agree and choice “3” indicated that students neither agreed nor disagreed. Table 20 highlights items where “neutral” responses provided incomplete information when trying to determine students’ experiences.

While the original IFOS instrument was normed with all categories separated, in this analysis, the “agree” and “strongly agree” categories were collapsed, and the “disagree” and “strongly disagree” categories were collapsed as that level of detail was not required for the purposes of this study.

Table 20*Treatment of “Neutral” Data When Considering Students’ Sensitivity to Feedback*

Question	<i>M</i> (<i>SD</i>)	Median	%Disagree	%Neutral	%Agree
1. My feelings can be easily hurt by corrective feedback from a teacher.	3.07 (1.06)	3	35	30	36
2. Corrective feedback is intimidating.	3.27 (1.06)	3	29	24	48
3. Corrective feedback is embarrassing.	2.92 (1.11)	3	36	31	34
4. I tend to dwell on the negative feelings that result from corrective feedback.	3.33 (1.03)	3	24	29	48

Note. Each response was on a five-point Likert-type scale where “strongly disagree” was “1” and “strongly agree” was “5.” The “1” and “2” scores were combined and scored “disagree,” the “3” score is neutral, and the “4” and “5” scores were combined and scored “agree” for this analysis.

Table 20 highlights several questions where the “neutral” responses provide incomplete information. In Question #1, “My feelings can be easily hurt by corrective feedback from a teacher,” it is possible that if the 30% of the students who chose “neither agree nor disagree” are grouped with the 36% of students who either agreed or strongly agreed, then 66% of students felt they were or could be hurt by feedback that teachers who participated in this study felt they were proficient or exemplary at providing. In Question #2, “Corrective feedback is intimidating,” if the 24% of students who chose “neither agree or disagree” are grouped with the 48% of students who either agreed or strongly agreed, then almost three quarters of participating students (72%) were intimidated by corrective feedback. The same holds true when

considering 65% of students who either chose the neutral response (31%) or agreed or strongly agreed (34%) that corrective feedback was embarrassing. Ryan and Henderson's (2018) research underscores the useful information that could be gained from getting more clarity about the 29% of participants who chose the neutral response on the Question #4, "I tend to dwell on the negative feelings that result from corrective feedback." If academic feedback is meant to provide students information for improvement, and if students are more likely to reject feedback that evokes negative emotional responses (Ryan & Henderson, 2018), then determining how to give feedback that motivates and inspires will have a direct impact on learning outcomes.

Limitations Due to Small Sample Size of Student Interviews

These data were limited by the fact that the sample size for the interviews included only 10 students. In addition, the participating students were 11th- and 12th-grade students only. It might be useful to compare findings from a 9th- and 10th-grade cohort with the findings from the 11th- and 12th-grade cohort to determine whether experience and maturity impact the results.

Limitations Due to Reliability and Validity of Teachers' Self-Reporting

The reliability and validity of these data might be limited because teachers were asked to rate their own skills with assessment practices, and teachers might not have an accurate perception of what equitable and effective practices are and/or their skill level with assessment practices.

Recommendations

These data indicate that there is a wide discrepancy between teachers' perceived skill at giving academic feedback and how students experience their feedback. While teachers in this study were asked to rate their skill levels in several areas of feedback and assessment practices, it was clear that they had little basis by which to gauge their performance. Few teachers reported having been provided instruction in their preservice programs, and most reported having no ongoing inservice related to assessment. The survey did not ask what specific programs teachers participated in during their credentialing process. Even when teachers reported having professional development focused on assessment, this experience did not necessarily mean they perceived themselves to be skilled. Therefore, school leaders should not make assumptions about the training and skill level of the faculty and, instead, they should provide comprehensive and ongoing training to ensure best practices. Along with participating in school-wide training opportunities, teachers could benefit from department and/or interdisciplinary professional learning communities where they examine practices and share experiences. Because this study revealed a gap between teachers' perceptions and students' experiences with assessments, teachers might conduct focus groups with students to learn more about how students experience instructional feedback and assessment practices. As noted in the literature review, female students experience feedback and assessment differently, and the impact on their academic self-efficacy and emotional health and well-being is greater. For these reasons, targeted professional development in feedback and

assessment practices might improve the academic success of female students, as well as their overall health and well-being. Specific recommendations include:

- Conducting research with student focus groups and asking for descriptions of their experiences with feedback;
- Having specific academic departments norm assessment practices by offering feedback and a grade on an anonymous student assignment;
- Offering training that highlights the importance of positive, timely, and specific feedback in the overall feedback and assessment process; and
- Providing training for both teachers and students on how to provide feedback and receive feedback; according to research, training students on how to receive feedback is a critical component of improvement that is often overlooked (Hattie & Timperley, 2007)

Implications for Future Research

To better understand how female students experience academic feedback and assessment, future research could be expanded to include all-girls schools in a variety of settings (urban and suburban and from different parts of the country) as well as female students from co-educational schools (private and public). Because the participating students in this study are all from a very specific and hard-to-replicate school setting, it would be valuable to compare the results of this study with a broader sample of female students.

While this study did not focus on girls' experiences with feedback and assessment within a specific discipline, it is interesting that participating students chose to bring only humanities-based assessments to discuss during their interviews.

This raises the question about whether feedback and assessment in STEM subjects have a similar impact on students. While there are studies focused on girls in STEM, a study that explores each female student's experiences when receiving feedback in various disciplines might help determine whether the impact of feedback is student-based, discipline-based, or a combination of both.

Conclusions

There is a saying attributed to Buddha and the tradition of Tibetan Buddhism: "When the student is ready, the teacher will appear." Philosophically, this concept demystifies the relationship between pupil and instructor, and is a beautiful notion of the natural way one looks to mentors in the desire to learn. But it could not be further from the reality of contemporary schooling. Whether or not students are ready, able, or interested, they are sent to school where professional educators are tasked with the responsibility of teaching content and skills, providing feedback for improvement, assessing students' progress, and reporting the results. Providing feedback and assessing learning are fundamental components of a teacher's role, and because teachers often lack adequate and timely training, they resort to practices that, at the very least, are inequitable and may even be harmful to students' motivation to learn, academic self-efficacy, and self-esteem (Feldman, 2019, 2020; Marques de Miranda et al., 2020).

Research shows that teachers consider assessment to be one of the least liked components of their job (Feldman, 2019; Guskey, 2011; Guskey & Bailey, 2001). While they feel less prepared to give effective feedback and grades than other parts of their job, it is a major part of their work. Many well-intentioned teachers may believe

themselves to be skilled at providing feedback, but the data from this study, albeit with a limited sample size, reveal that teachers might not be as skilled as they believe. In fact, as shown in this study, there is often a considerable gap between teachers' level of training in assessment and their perceived skill level. While this disparity may cause frustration for teachers, it might be harmful to students' motivation to learn, their academic self-efficacy, and their overall health and well-being (Brookhart, 2013; Feldman, 2019; Villeneuve et al., 2019). And the literature shows that these harmful outcomes are more pronounced for female students (Altermatt & Pomerantz, 2003; Johnson & Helgeson, 2002). Research suggests that compared to their male counterparts, female students experience higher levels of academic anxiety partly in response to the way they process teacher feedback. This study also reveals that there is a strong emotional connection to their schoolwork, especially their writing. When the female students in this study experienced positive, specific, and timely feedback from a teacher, they felt "proud" and that "feedback [is] the best parts of an education and it's what really helps with growth." When the female students in this study experienced negative feedback, they reported feeling "scared," and one student even had to "immediately go to the counselor's office because [she] wanted to kill [herself]." While this is an extreme response to a student receiving a "C" on an exam, teachers need to be aware that feedback and grades could trigger an extreme response in a fragile student.

Teaching necessitates giving corrective feedback, and in giving feedback there is always a risk that students will react negatively or even with an extreme response. Simply reminding teachers that assessment practices and outcomes are problematic

does little to alleviate their frustration or narrow the gap between what teachers believe they are doing and what they actually are doing. This research calls for targeted and differentiated professional development to improve feedback and assessment practices and furthering the understanding of the ways students experience instructional feedback. This training is not only critical to teaching and learning, but it may also profoundly impact the health and wellness of our students, especially our female students.

References

- AFT, NCME, & NEA. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: National Council on Measurement in Education.
- Alivernini, F., & Lucindi, F. (2011). Relationship between social context, self-efficacy, motivation, academic achievement, and intention to drop out of high schools: A longitudinal study. *The Journal of Educational Research, 104*, 241-252.
- Altermatt, E. R., & Pomerantz, E. M. (2003). Why girls worry about school performance. *Principal, 82*(4), 67-68.
- Altermatt, E. R. (2015). Predicting day-to-day changes in students' school-related affect from daily academic experiences and social interactions. *The Journal of Experimental Education, 83*(2), 218-235.
- Anand, S. & Bhatia, T. (2021). COVID-19 pandemic: Challenges among rising high school seniors, class of 2021 and effect on stress regarding college admissions. *Journal of Student Research, 10*(1), 1-14.
- Anxiety. American Psychological Association. Retrieved from <https://www.apa.org/topics/anxiety/>
- Bandura, A. (1997). *The exercise of control*. New York: W. H. Freeman.
- Battistone, W., Buckmiller, T., & Peters, R. (2019). Assessing assessment literacy: Are new teachers prepared to assume jobs in school districts engaging in grading and assessment reform efforts. *Studies in Educational Evaluation, 62*, 10-17.

- Bian, L., Leslie, S.-J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist*, 73(9), 1139-1153.
- Biernat, M., & Danaher, K. (2011). Interpreting and reacting to feedback in stereotype-relevant performance domains. *Journal of Experimental Social Psychology*, 48, 271-276.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575.
- Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. *International Encyclopedia of Education*, 7, 548-556.
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Brookhart, S. M. (2013). The use of teacher judgment for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69-90.
- Bryk, A., Lee, V., and Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.
- Burry-Stock, J. A., & Frazier, C. H. (2005). *Assessment practices inventory (Revised)*, Unpublished manuscript, Tuscaloosa, Alabama.

- Costello, E. J., Erkanli, A., & Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry*, *47*(12), 1263-1271.
- Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, *40*(1), 67-87.
- Craig, T. A. (2011). *Effects of standards-based report cards on student learning*. Boston, MA: Northeastern University.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches (4th ed.)*. Thousand Oaks, CA: Sage.
- Duckworth, A. L., & Seligman, E. P. (2006). Self-discipline gives girls the edge: gender in self-discipline grades, and achievement tests scores. *Journal of Educational Psychology*, *98*(1), 198-208.
- Durm, M. W. (1993, Spring). "An A is not an A: A history of grading." *The Educational Forum*, *57*, 294-297.
- Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter?: Evidence from NELS88, National Bureau of Economic Research (NBER Working Paper No. 4669). Retrieved from <http://www.nber.org/papers/w4669>
- Feldman, J. (2019). *Grading for Equity*. Thousand Oaks, CA: Corwin Press, Inc.
- Feldman, J. (2020). Taking the stress out of grading. *Educational Leadership*, *78*(1), 14-20.
- Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational Research: Competencies for analysis and applications*. (10th ed.). Boston: Pearson Education, Inc.

- Green, S., Johnson, R. L., Kim, D. H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23, 999-1011.
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communications and Technology Journal*, 30(4), 233-252.
- Guskey, T. R. (1996). *Reporting on student learning: Lessons from the past—prescriptions for the future* (pp 13-24). In T. Guskey (Ed.) *ASCD Yearbook 1996: Communicating student learning* (pp 141-177). Alexandria, VA: Association for Supervision and Curriculum Development.
- Guskey, T. R. (2007). Multiple sources of evidence: an analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice*, 26(1), 19-27.
- Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership*, 69(3), 16-21.
- Guskey, T. R. & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Gutek, G. L. (1986). *Education in the United States: An historical perspective*. Englewood Cliffs, NJ: Prentice-Hall.
- Haertel, E., & Herman, J. (2005). A historical perspective on validity arguments for accountability testing. *Yearbook of the National Society for the Study of Education*, 104(2), 1-34.

- Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 34*(3), 269-290.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *PHI DELTA KAPPAN, 89*, 140-145.
- Imperial, P. (2012). Grading and reporting in Catholic high schools: Problems and promise. *Momentum, 43*(3), 20-24.
- Howell, C. A. (2013). *Development and analysis of a measurement scale for teacher assessment literacy* [Unpublished master's thesis]. East Carolina University.
- Johnson, M. & Helgeson, V. S. (2002). Sex differences in response to evaluative feedback: A field study. *Psychology of Women Quarterly, 26*, 242-251.
- King, Paul E., Schrodt, Paul, & Weisel, Jessica J. (2009). The instructional feedback orientation scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. *Communication Education, 58*(2), 235-261.
- Lau, F., Antonio, M., Davison, K., Queen, R., & Devor, A. (2020, November). A rapid review of gender, sex, and sexual orientation documentation in electronic health records. *Journal of the American Medical Informatics Association, 27*(11), 1774–1783. <https://doi.org/10.1093/jamia/ocaa158>
- Lemann, N. (1997). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus, and Giroux.

- Madigan, D. J. (2019). A meta-analysis of perfectionism and academic achievement. *Educational Psychology Review, 31*(2), 1-23.
- Mael, F., Alonso, A., Gibson, D. Rogers, K., & Smith, M. (2005). *Single-sex versus coeducational Schooling: A systematic review*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- Magson, N., Freeman, J., Rapee, R., Richardson, C. E., Oar, E., & Fardouly, J. (2020). Risk and protective factors for prospective changes in adolescent mental health during the COVID-19 pandemic. *Journal of Youth and Adolescence, 50*, 44-57.
- Mahmood, D. & Jacobo, H. (2019). Grading for growth: Using sliding scale rubrics to motivate struggling learners. *Interdisciplinary Journal of Problem-Based Learning, 13*(2).
- Major, B. & Schmader, T. (1998). Coping with stigma through psychological disengagement. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 219-241). San Diego, CA: Academic Press.
- Marques de Miranda, Athanasio, B. S., Sena Oliveira, A. C., & Simoes-e-Silva, A. C. (2020, December). How is COVID-19 pandemic impacting mental health of children and adolescents? *International Journal of Disaster Risk Reduction, 51*, p.101845.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2010). *Formative assessment and standards-based grading*. Bloomington, IN: Marzano Research.

- Mazure, C. (2021, September 19). *What do we mean by sex and gender?* Yale School of Medicine. <https://medicine.yale.edu/news-article/what-do-we-mean-by-sex-and-gender/>
- McDermott, E. J. (1997). *Distinctive qualities of the Catholic school*. Washington, DC: National Catholic Educational Association.
- Mechtenberg, L. (2009). Cheap talk in the classroom: How biased grading at school explains gender differences in achievements, career choices and wages. *Review of Economic Studies*, 76, 1431-1459.
- Middleton, W. (1933). Some general trends in grading procedure. *Education*, 54(1), 5-10.
- Mohanty, N. & Jena, S. (2015). Understanding mental health of adolescents. *Journal of Health and Wellbeing*, 6(10), 968-972.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of Students*. (6th ed.). Boston: Allyn & Bacon.
- O'Connor, K., Jung, L. A., & Reeves, D. (2018). Gearing up for FAST grading and reporting. *Phi Delta Kappan*, 99(8), 67-71.
- Pew Research Center. (2015). *Women and leadership: Public says women are equally qualified, but barriers persist*. Washington, DC. <https://pewrsr.ch/2niKcdO>
- Pew Research Center. (2019). *Most U.S. teens see anxiety and depression as a major problem among their peers*. Washington, DC. <https://www.pewresearch.org/social-trends/2019/02/20/most-u-s-teens-see-anxiety-and-depression-as-a-major-problem-among-their-peers/>

- Pomerantz, E. V., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94*(2), 396-404.
- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education, 33*(2), 143-154.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380.
- Rinfret, N., Tougas, F., Beaton, A. M., Laplante, J., Manguelle, C. N., & Lagace, M.C. (2014). The long and winding road: grades, psychological disengagement, and motivation among female students in (non)-traditional career paths. *Social Psychology of Education, 17*, 637-656.
- Roberts, T.- A. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin, 109*(2), 297-308.
- Rosenberg, F. R., & Simmons, R. G. (1975). Sex differences in self-concept in adolescence. *Sex Roles 1*, 147-159.
- Ryan, T., & Henderson, M. (2018). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education, 43*(6), 880-892.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd Ed.). Thousand Oaks, CA: Sage Publications.
- Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., & De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. *Scientific Reports, 10*(11112). <https://doi.org/10.1038/s41598-020-68008-z>

- Sawchuck, S. (2020, April 1). *Grading students during the coronavirus crises: What's the right call?* Retrieved from <https://www.edweek.org/teaching-learning/grading-students-during-the-coronavirus-crisis-whats-the-right-call/2020/04>
- Schwarz, R., Robinson, M., Kirst, M., & Kirp, D. (2000). Goals 2000 and the standards movement. *Brookings Papers on Education Policy*, 3, 173-214.
- Schunk, D. H., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85-104). New York: The Guilford Press.
- Spencer, K. (2012). Standards-Based Grading. *The Education Digest* 78(3), 4-10.
- Scott, S. R., Rivera, K. M., Rushing, E., Manczak, E. M., Rozek, C. S., & Doom, J. R. (2021). I hate this: A qualitative analysis of adolescents' self-reported challenges during the COVID-19 pandemic. *Journal of Adolescent Health*, 68, 262-269.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263-300.
- Shin, J., Lee, Y., & Seo, E. (2017). The effects of feedback on students' achievement goals: Interaction between reference of comparison and regulatory focus. *Learning and Instruction*, 49, 21-31.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.

- Staats, C. (2014). *Implicit racial bias and school discipline disparities* (Kirwan Institute Special Report). Columbus: Ohio State University.
- Starch, D. (1916). *Educational Measurements*. New York. The Macmillan Company.
- Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high school work in English. *School Review*, *20*, 442-457.
- Starch, D., & Elliott, E. C. (1913). Reliability of the grading of high school work in Mathematics. *School Review*, *21*. 254-259.
- Stephens-Davidowitz, S. (2014, January 18). Google, tell me. Is my son a genius? *The New York Times*. Retrieved from <https://nyti.ms/1azD4xD>
- Storage, D., Horne, Z., Cimpian, A., & Leslie, S. J. (2016). The frequency of “brilliant” and “genius” in teaching evaluations predicts the presentation of women and African Americans across fields. *PLoS ONE*, *11*, e0150194. <http://dx.doi.org/10.1371/journal.pone.0150194>
- Tierney, R. D. (2015). Altered Grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, *8*, 5-30.
- Twenge, J. M., Cooper, A. B., Joiner, T. E., & Binau, S. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, *128*(3), 185-199.
- United States Conference of Catholic Bishops. (2002). *Renewing our commitment to Catholic elementary and secondary schools in the third millennium*. Washington, DC: United States Conference of Catholic Bishops.

- Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based learning*. Alexandria, VA: ASCD.
- Villeneuve, J. C., Conner, J. O., Selby, S., & Pope, D. C. (2019). Easing the stress at pressure-cooker schools. *Phi Delta Kappan*, 101(3), 15-19.
- Wilkinson, D., & Birmingham, P. (2003). *Using research instruments: A guide for researchers*. Psychology Press.
- Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report*, 20(2), 34-152.
- Yin, R. K. (2018). *Case study research and applications: Design and methods*. (6th ed.). Thousand Oaks, CA: Sage.

Appendix A

Scale of Teacher Assessment Practices (STAP)* Instructions

The following survey measures multiple aspects of teacher assessment practices in the classroom. The data is being collected for two purposes:

1. To assist the school in determining opportunities to provide additional training for teachers on assessment; and
2. To add to data being collected for a study on how female students experience academic feedback and assessment.

Thank you for participating in this survey, and please know that your responses will remain confidential.

Directions: There are 36 items relating to classroom assessment practices. Each item is followed by a scale ranging from “1” (very low) to “5” (very high). Please estimate the level of your skills with each practice and circle the appropriate number. For example, if you feel that your skills are “Very High” with regard to “Explaining assessment results clearly to parents” then you would circle number “5”. In contrast, if you feel that your skills relating to this assessment practice are “Very Low” then you would circle number 1. If you do not engage in a particular practice, please choose “n/a”. The second section of the survey asks about demographic information.

*This instrument was used with permission of the copyright owner Catherine Howell 2013. It was based on questions that were reproduced with permission of the copyright owners. Copyright Judith A Burry-Stock and Celeste H. Frazier Assessment Practices Inventory (Revised) (APIR) 2005.

STAP Survey (Qualtrics):

	Very Low	Low	Acceptable	High	Very High	n/a
1. Explaining assessment results clearly to parents	1	2	3	4	5	n/a
2. Seeking assistance when I am unsure how to score an item	1	2	3	4	5	n/a
3. Choosing an assessment method for a specific purpose relating to an individual student	1	2	3	4	5	n/a
4. Using the results of formative assessment to adjust the content of my lessons	1	2	3	4	5	n/a
5. Adhering to the bounds of confidentiality regarding assessment results	1	2	3	4	5	n/a

6. Using assessment results to appropriately group students for instruction	1	2	3	4	5	n/a
7. Selecting appropriate methods for reporting results to others in addition to grades	1	2	3	4	5	n/a
8. Explaining results to other educators for the purpose of assisting with placement decisions	1	2	3	4	5	n/a
9. Using results of summative assessments to adjust future lesson plans	1	2	3	4	5	n/a
10. Knowledge of the consequences of unethical use of assessment	1	2	3	4	5	n/a
11. Selecting multiple methods of assessment (e.g. tests observations)	1	2	3	4	5	n/a
12. Recognizing inappropriate use of assessment	1	2	3	4	5	n/a
13. Interpreting summary scores reported with standardized test results (e.g. mean, percentile rank)	1	2	3	4	5	n/a
14. Administering progress-monitoring assessments	1	2	3	4	5	n/a
15. Creating assessments that accommodate the needs of a variety of students	1	2	3	4	5	n/a
16. Explaining to parents how assessment results are used to make decisions about their children	1	2	3	4	5	n/a
17. Determining if an assessment is aligned with required standards (e.g. state or district curriculum goals)	1	2	3	4	5	n/a
18. Knowledge of which externally produced assessments are current and available	1	2	3	4	5	n/a
19. Administering standardized assessments (e.g. standardized achievement tests)	1	2	3	4	5	n/a
20. Recognizing when assessment results are being used inappropriately by others	1	2	3	4	5	n/a
21. Communicating the results of assessments to students in a way that they can understand	1	2	3	4	5	n/a
22. Using assessment information to develop an instructional plan for a student	1	2	3	4	5	n/a

23. Using progress monitoring results to adjust instruction	1	2	3	4	5	n/a
24. Using assessment results to identify students with similar needs	1	2	3	4	5	n/a
25. Interpreting criterion-referenced scores	1	2	3	4	5	n/a
26. Understanding why standardized administration is necessary to interpret results of standardized tests	1	2	3	4	5	n/a
27. Developing assessments with different formats (e.g. multiple-choice fill-in-blank short answer)	1	2	3	4	5	n/a
28. Identifying my own legal responsibilities in regard to assessment	1	2	3	4	5	n/a
29. Sampling from the domain defined by learning goals to write assessment items	1	2	3	4	5	n/a
30. Explaining to students how assessment results will be used to assign grades	1	2	3	4	5	n/a
31. Writing fill-in-the-blank/short answer questions	1	2	3	4	5	n/a
32. Using assessment results when developing lesson plans	1	2	3	4	5	n/a
33. Revising a test based on item analysis	1	2	3	4	5	n/a
34. Using assessments such as classwork to enhance my instructional delivery	1	2	3	4	5	n/a
35. Using assessment results to improve teaching and learning	1	2	3	4	5	n/a
36. Developing assessments based on clearly defined course objectives	1	2	3	4	5	n/a

Demographic Information:

1. Age
2. Gender
3. Ethnicity
4. Number of years teaching
5. Grade level(s) currently taught
6. Have you ever taken a course in assessment? Yes No
If yes, how many?
7. Have you ever attended an inservice on assessment? Yes No
If yes, how many?

8. I have taken a course within the past five years specifically designed for equitable assessment practices? Yes No
9. My grades include a combination of demonstrated academic achievement and non-academic elements, such as effort, participation, and/or attendance. Yes No
10. I would like additional inservice opportunities related to equitable assessment. Yes No

Appendix B

Student Survey (Qualtrics)

Survey: We want to better understand how students experience grades and academic feedback (comments from teachers on written work, projects, presentations, and so forth) at our school. For the purposes of this survey, academic feedback includes written or verbal comments about the quality of your work or performance and/or comments meant to improve your work or performance. Your answers will remain confidential. The second part of the survey asks about demographic information. Thank you for taking part in this survey. The information collected is part of a research project for a doctoral student at the University of Portland. It will also be used to help school administration identify themes and patterns regarding students' experiences.

Instructions: Please answer the following questions using the five-point scale where (1) is strongly disagree and (5) is strongly agree.

Question	Strongly Disagree 1	Disagree 2	Neither Agree/ Disagree 3	Agree 4	Strongly Agree 5
Utility (headings will not be included)					
1. I think feedback from teachers is vitally important in improving my performance.	1	2	3	4	5
2. I will usually reflect on a teacher's feedback.	1	2	3	4	5
3. I listen carefully when a teacher provides feedback.	1	2	3	4	5
4. I am extremely encouraged by positive feedback from teachers.	1	2	3	4	5
5. I think that feedback provides clear direction on how to improve my performance.	1	2	3	4	5
6. Feedback from my teachers can be a valuable form of praise.	1	2	3	4	5
7. I pay careful attention to instructional feedback.	1	2	3	4	5
8. Feedback from my teachers motivates me to improve my performance.	1	2	3	4	5
9. Feedback from teachers is a waste of time.*	1	2	3	4	5

10. I feel relieved when I receive positive feedback.	1	2	3	4	5
Sensitivity					
11. My feelings can be easily hurt by corrective feedback from a teacher.	1	2	3	4	5
12. I feel threatened by corrective feedback.	1	2	3	4	5
13. Corrective feedback hurts my feelings.	1	2	3	4	5
14. Corrective feedback is intimidating.	1	2	3	4	5
15. My feelings are not easily hurt by corrective feedback from a teacher.*	1	2	3	4	5
16. It is difficult to “get over” corrective feedback.	1	2	3	4	5
17. Corrective feedback is embarrassing.	1	2	3	4	5
18. I tend to dwell on the negative feelings that result from corrective feedback.	1	2	3	4	5
19. Corrective feedback from a teacher increases the stress I feel about future performance.	1	2	3	4	5
Confidentiality					
20. I do not like to receive corrective feedback in front of other people.	1	2	3	4	5
21. I do not like for others to hear what feedback I am receiving.	1	2	3	4	5
22. I don’t mind being singled out by feedback from a teacher.*	1	2	3	4	5
23. I prefer to receive feedback from a teacher in private.	1	2	3	4	5
24. I like others to hear the feedback I am receiving from my teacher.*	1	2	3	4	5
Retention					
25. I can’t remember what teachers want me to do when they provide feedback.	1	2	3	4	5
26. I tend to miss out on the details of what instructors want when	1	2	3	4	5

they provide me with feedback.					
27. I typically do not make note of the teacher's corrective comments.	1	2	3	4	5

Note: Instructional Feedback Orientation Scale has been used with permission (King et al., 2009).

*These items are reverse coded.

Demographic Information:

Please answer the following information

Grade: 11 12

Ethnicity:

Asian/Pacific Islander

Black

Hispanic

Middle Eastern

Multi-Racial

Native American

White

Other not listed

Prefer not to answer

Would you describe yourself as:

an extravert/outgoing

an introvert/more reserved

Appendix C

Student Interview Protocol

Hi _____, How are you doing today? How is your year going so far?

Thank you for your willingness to participate in this follow-up interview. As you know from the student survey, Ms. Foran is studying at the University of Portland and is collecting data for a research project to try to better understand how students experience teachers' feedback and grades, which for this interview includes written comments and final grades.

Thank you for bringing a graded assignment with you. If you have more than one, please choose the one that you might have the most to say about. I have a set of questions to ask you, and I might also ask that you expand on an answer or give more details. Please know that your teachers will not have access to your personal information and will not be able to identify you in any way. I really appreciate the opportunity to talk with you today, so let's get started.

Interview Questions:

1. Please describe the assignment you've brought with you. Response should include the following (follow up with questions to get a complete description of assignment):
 - Class
 - Description of the original assignment guidelines or unit/text/lab if assessment
 - When in the school year did you complete? (fall, winter, spring)
 - Was this a unit project, part of a unit?
2. Talk with me about why you chose this assignment/test to talk about today.
3. Was there feedback on this that you found especially helpful or encouraging?

Follow-up:

- Tell me how it helped you. OR
 - Did you use this feedback as you were completing other assignments? (for this class? for other classes?)
4. Was there feedback that you found discouraging or unfair/unjust? (can substitute insensitive if already covered)?

Follow-up:

- Please share what, specifically, felt unfair.
- Did it impact how you approached other assignments for this class? (or for another class?)
- If so, in what way?

5. What were your key takeaways from the feedback and/or final grade on this piece?

Follow-up:

- Did this feedback motivate you to work harder/spend longer studying/completing your work?
6. Does the feedback on your work influence the way you view yourself as a student? Can you explain? (looking for academic self-efficacy/identity/motivation/self-confidence here)
7. Thinking about assessments or feedback you have received from teachers in high school, please describe any that you remember as particularly encouraging or discouraging and why?

Appendix D

Directions for choosing assignment(s) to bring to interview: Choose one or two graded assignments to talk about during the interview. This can be a project, test, essay, lab, or any longer assignment of your choice. Nightly math homework or vocabulary quizzes are examples of graded work that would not be good choices. As you consider your choices, you might look for graded work that meets the above criteria, but does not have enough feedback to give you the information you think might better support the grade you received. Please reach out to me (this will be the third party) in advance of our scheduled interview if you have questions.

Appendix E

GUARDIAN AUTHORIZATION:

Dear St. Mary's Parents of Juniors and Seniors:

I am currently working on a doctoral degree at the University of Portland and the topic of my thesis is learning more about how students experience academic feedback and assessment. I am inviting all 11th and 12th grade students to participate in a survey as part of my research study. If you decide to allow your student to participate, she will be asked to complete an electronic survey in her homeroom class that takes approximately 10-15 minutes to complete. She will not be asked to give her name on this survey, and her participation is voluntary. Your decision whether or not to allow your student to participate will not affect your or your student's relationship with St. Mary's Academy. If you decide to allow your student to participate, you and/or your student are free to withdraw your consent and discontinue participation at any time.

Your student's identity will remain confidential, and Ellie Gilbert, the school's librarian, will conduct the surveys and compile the data to further ensure confidentiality. If you have any questions about the study, please feel free to contact me at nicole.foran@smapdx.org or my University of Portland Dissertation Chair, Dr. Randy Hetherington at hetherin@up.edu. If you have questions regarding your student's rights as a research subject, please contact the IRB (irb@up.edu).

Please click on this link ([link here](#)) to give your student permission to participate in this study. Your electronic signature on the permission form indicates that you have read and understand the information provided above, that you willingly agree to allow your student to participate, that you and/or your student may withdraw your consent at any time.

This research is an important component in helping me learn more about students' experiences and I appreciate your consideration in allowing your student to participate.

Sincerely,
Nicole Foran