2015

Middle School Students' Perceptions of Fairness and Trust in Assessment Scenarios

Corinne Elaine Jimenez
University of South Carolina

Recommended Citation

Jimenez, C. E.(2015). *Middle School Students' Perceptions of Fairness and Trust in Assessment Scenarios.* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/3677

Middle School Students' Perceptions of Fairness and Trust in
Assessment Scenarios

By Corinne Elaine Jimenez

Bachelor of Arts
State University of New York at Binghamton, 1994

Master of Arts in Teaching
University of South Carolina, 1996

Master of Education
University of South Carolina, 2000

_____

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Educational Psychology and Research

College of Education

University of South Carolina

2015

Accepted by:

Robert L. Johnson, Major Professor

Xiaofeng Liu, Committee Member

Tammie Dickenson, Committee Member

John Arnold, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Dedication

To the two people who hold my heart and

made pursuing my dreams possible.

I nub new, Fernando and Alex Jimenez

Acknowledgements

I was the first in my mother's family to graduate college. I am the first in her family to earn a doctorate. But neither would be possible without my mother's consistent expectation of nothing less than greatness. Thank you for being my first teacher and forever cheerleader.

My father, to whom I am a princess, taught me life isn't always fair, but you move past it. Maybe that is why I chose a topic that explores students' sense of fairness, to help make life just a little more fair and to honor him.

Margaret Hicks has been my friend and colleague for at least a decade. Without her sending me articles and pushing me professionally, I would still be deciding on a topic. She has been an editor, a sounding board, and a shoulder to lean on throughout the process.

Robert Johnson is the reason I am a doctor of philosophy. His unwavering support of my passion to improve classroom assessment practices has been instrumental in my academic and professional career.

John Arnold has been the best colleague any person could ask for. I have been able to share with him my frustrations and worries, and he is always there with a kind word and wonderful smile.

Abstract

Fairness in assessment practices is an elusive concept which has been explored in some detail at the university level, but rarely with students in middle school. This study examines students' perceptions of fairness on nine classroom assessment practices. It also studies if students' perceptions of fairness, as well as student gender, could predict their levels of trust. Students were able to discern fair practices, as defined by alignment to best assessment practice literature, from unfair practices for most scenarios. They were more inclined to recognize interactional fairness deviations or promotions. The student's ability to identify a fair or unfair situation did predict the level of trust they would have in a teacher, though the student's gender did not. Recommendations include teachers being more cognizant of their assessment practices and how they will be seen by their students in terms of fairness.

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

Fairness is an elusive concept. Yet, for most, it is an ideal that guides their interactions with individuals, groups, and institutions (Tyler, Boekmann, Smith, & Huo, 1997). Students have an opportunity to experience fairness, or unfairness, daily. In one day, a teacher, group of peers, and school policy might be judged based on a student's sense of fairness. Vieno, Perkins, Smith, & Santinello (2005) found that schools in which students felt the teachers acted fairly and the rules were applied fairly had a higher sense of community.

Standardized testing often comes to mind for the general public in terms of fairness. The concept behind standardized tests is to allow all students a similar opportunity to show their achievement on a given set of objectives. However, due to issues with accessibility to the test or content and universal design of the test, sometimes these standardized examinations are considered unfair (American Educational Research Association, 2014). In the most recent *Standards for Educational and Psychological Testing*, the American Educational Research Association devotes an entire chapter to fairness in standardized testing, calling fairness an "overriding, foundational concern" (p. 49). The results of these standardized tests are then used for various decisions: retention, tracking, and graduation to name a few (McNeil, 2002). Because these decisions

can make a dramatic impact on a student's future, research studies began to look at the ethics of teacher actions in terms of standardized test preparation (Mehrens & Kaminski, 1989).

However, standardized tests are only given at limited times in a students' career. Classroom assessment and grading often are considered when one speaks of unfair actions at the school level. Also, assessment and evaluation are of great importance when designing curriculum. Martone and Sireci (2009) studied assessment as part of curricular design and found an important component of effectiveness was the "teachers need to understand the value of the assessments, how the assessments relate to what they should be teaching, and how to make changes in their approach based on the results they see" (p. 1356).

Assessment is not always a separate part of the daily lesson. Tierney (2014) described formative assessment as the feedback teachers give to students as the teacher is instructing and questioning, with the results of those interactions being used to inform instruction. Tierney goes on to explain summative assessment is used to "report on achievement" (p. 55).

Assessment, be it formative or summative, is the means to judge how well a student understands the taught concepts. Thus, effective teachers must assess well, but they must also be fair in their assessment practices. The use of best practices in assessment and grading may be one way to address this issue of fairness.

Many experts have studied the concept of best practices in assessment and grading. However, as standards-based instruction has become the norm in the United States, assessment practices have not quite kept up. As Guskey (2009) stated, "of all aspects of our educational system, none seems more impervious to change than grading" (p. 2). Various researchers have recommended assessment policies including not using zeros in grading (Carifo & Carey, 2009; Guskey, 2009; Wormeli, 2006) and only using a student's demonstrated ability, as opposed to effort and growth, in reporting achievement (Edwards, 2000; Guskey, 2009; Stanley & Baines, 2004). Yet, teacher's individual grading practices widely vary and often stray from the best practice assessment literature. For example, Tierney, Simon and Charland (2011) found teachers will often knowingly abandon their pedagogical knowledge in order to maintain, what they esteem to be, a fair grading environment. However, these "fair" practices "may have included a host of unexamined assumptions" (p. 224), such as assessing students in a lower track on a set of non-grade level content standards merely because they were in the low group, not because of individual need or performance. This form of bias undermines a key aspect of fairness (McMillan, 2004).   This is one example of the confusion that surrounds the claim by teachers and administrators of fair and accurate assessments of student achievement.

This line of reasoning leads back to the initial concept that fairness is an elusive ideal. Teachers' perceptions of what constitutes a fair grading practice

vary.   However, teachers are just one part of the educational equation. The students who educators serve are the clients. One might argue, if students find the assessments and subsequent grading practices to be unfair, their purpose of providing valuable and accurate feedback to the student for further growth might become obsolete. In describing attribution theory, Carifio and Carey (2009) explained "grading practices must be perceived as reliable and predictable if they are to have the intended and desired effect" (p. 29). Hence, it might follow that if students do not find the teacher's assessment policy and procedures to be fair because they are not reliable nor predictable, they might not perform in a true form. In terms of classical test theory, the observed score on the test or project might be heavily influenced not by the student's true score, but by the error associated with the student's sense of unfairness. Thus, assessments would not give the teacher a realistic picture of what the student knows, and would not allow for meaningful reflection on how to change instruction to meet the needs of the child. This logical progression might apply to all stages of the assessment and grading procedures: from formative to summative assessment, and in terms of overall course or subject evaluation.

Trust is another elusive concept which might relate to fairness. If trust is the "firm belief in the reliability, truth, ability or strength of someone or something" (Oxford Dictionaries, 2015), it would seem a teacher who acts in a fair manner, a reliable and truthful manner, would foster trust in his or her students.

Some research has been conducted on the importance of trust in schools. In their report on school climate, Thapa, Cohen, Hiffins-D'Allessando, and Guffey (2012) found "a positive school climate promotes cooperative learning, group cohesion, respect, and mutual trust" (p. 7). These researchers discussed many forms of mutual trust: parent-teacher, teacher-administrator, school level-district-level, teacher-teacher, and more. However, specific practices which build or derailed trust were not investigated, nor was the relationship between fairness and trust.

Bryk & Schneider (2004) used case studies to build a four system model for an effective school. They believed professional capacity, order and safety, parent-school ties, and instructional guidance were all essential and were all tied together by the various stakeholders trusting each other. While they did report fair discipline and classroom management practices were important to creating and maintaining trust, they did not report on any relationship between fairness and trust in terms of assessment practices.

How do students view different assessment practices in terms of fairness? Does their perception of fairness impact their trust in the teacher? In the past, researchers have studied the perceptions of college students and their professors' grading policies and feedback mechanisms (Gordon & Fay, 2010; Holmes & Smith, 2005). Other researchers have investigated middle school students' preferences for different assessment techniques (James, Griffin, & Dodds, 2009; Wang & Holcombe, 2010). However, reports into middle school

students' insights about the fairness of different common assessment practices are limited.

To understand more about how students view assessment and grading in terms of fairness, the following research questions were investigated:

Do students' perceptions of fairness in an assessment scenario differ based on the characteristics of the scenario? More precisely, do students perceive assessment practices that promote best practice as fair, and those that deviate from best practice as unfair?

Then, in order to examine the possible relationship between fairness and trust, the following secondary research question was asked:

To what extent does a student's perception of fairness and the student's gender predict his or her perception of change in trust toward the teacher? In other words, if a teacher acts in a perceived fair way, does the student's trust in that teacher increase? Or, if a teacher acts in a perceived unfair manner, does the student's trust in that teacher decrease? Does student gender aid in this prediction?

To answer these research questions in the proposed study, different fictional scenarios were developed, some of which were written in accordance to best assessment and grading practice and some of which deviate from said practice. These scenarios also represent different types of fairness, which will be described more thoroughly in Chapter 2. After receiving feedback on the

scenarios from middle school students, the study examined the differences in perceived fairness scores based on the situation's fairness type.

In order to study the second research question, this investigation examined the relationship between students' views on the fairness of a scenario, student gender, and the change in trust they would have if their teacher should act in such a way as described in the setting. In other words, to what extent did a student's perception of fairness of an assessment circumstance and a student's gender predict his or her perception of change in trust toward the teacher that circumstance would bring about?

In addressing these questions, the study provided information about how middle school students perceived different assessment and grading scenarios in terms of fairness. It also offered insight into how fairness and trust are related. Study findings might suggest teachers need to consider student views on common assessment and grading practices, especially if those practices are found to be considered unfair and degrade the trust a student has in the teacher.

Hypotheses

There were two foci of this study. First, the study examined the differences between fairness scores based on characteristics of the assessment scenario. Secondly, it investigated the possible relationship between fairness scores, gender, and change in teacher trust of student. The guiding theoretical framework for the study was grounded in the assumption that a student's sense

of a teacher's fairness can be changed by a teacher's assessment practices, and that when a teacher acts unfairly, trust in that teacher is diminished. This model also assumed that the best practices of assessment, as described by various assessment experts, are fair. Hence, if a teacher does not follow best practice, that instructor is acting in an unfair manner. Therefore, one working hypothesis was there will be a difference in fairness perception scores, based on the underlying best practice characteristic of the scenario. The second working hypothesis was that as a fairness perception score increases, so will the student's trust in a teacher. There will be a strong, positive relationship between perception of fairness and change of trust in teacher. This hypothesis stemmed from the assumption that students trust a teacher based on his or her actions, including actions involving grading and assessments. Actions that are considered unfair will derail that trust; actions that are fair will build that trust. Also, gender will not be a significant regression factor in predicting a student's change in trust.

Generalization and Limitations

These findings might not be appropriate to generalize to elementary or high school situations given the unique transitional nature of traditional middle school. Middle school students change greatly from their sixth grade to their eighth grade years, becoming less immature and more like young adults. Also, the developmental level of elementary school students might limit the applicability of the findings, as they are less likely to question a teacher's judgment. Students in high school, on the other hand, move into a

8

developmental stage that demands more independent thinking, which results in questioning most authority figure's judgment, not based on merit but based on the fact that it came from an authority figure (Christine & Viner, 2005).

These data are from one suburban school district with an average (mid-point) poverty indicator (49% free and reduced lunch). However, this district is in the 93rd percentile in terms of wealth, as measured by the state's poverty index (South Carolina Department of Education, 2014). Hence, care should be taken when attempting to generalize these data to other settings.

Given the large sample size and use of an instrument based on widely accepted best assessment practices, the research might be generalized to other similar populations. It is also reasonable that the findings might be generalized to other subjects not covered in the scenarios, such as physical education and art, as the assessment best practices could be imagined in every type of class setting.

Chapter Two

Literature Review

The following sections provide a review of the literature for fairness in terms of the educational environment; the relationship between fairness, trust, and teacher effectiveness; and the potential for differences between teacher and student perception. Fairness will be described in relationship to ethics, as a construct, as perceived in a classroom, and as viewed through educational assessment measurements. The link between fairness and trust will be explored, as well as the connection between trust and teacher effectiveness. Finally, perception itself and the nature of perception being different based on one's role will be discussed.

Introduction

A teacher's responsibility, at the macro level, is to help create a more informed student population which will lead to a better citizenry. However, at the micro level, a teacher's purpose is to serve the individual student (Sloan, 2012). To serve a student, a teacher must be competent. One quality that a teacher must possess to be considered competent is credibility (Brookfield, 2006). Students must believe in the teacher and trust the teacher, in content and action. Whitley, Perkins, Balogh, Keith-Spiegel, and Wittig (2000) reasoned that a student's perception of a teacher's fairness can impact the trust between

the student and teacher. Teachers who are perceived as less fair can be trusted

less by students. This reasoning seems sound; however, the assessment

literature contains little research into actual student perceptions of fairness.

To follow the reasoning of Brookfield in terms of credibility (2006), this

study first examined if students feel a practice is fair or unfair. Then, it

measured to what extent the students' perception of the fairness of a practice

affects the trust the student has for the teacher. In Rodabaugh's (1996)

research on fairness, she identified actions which could be considered unfair by

college level students. Most of these actions are associated with either formative

or summative assessment. Examples of fair practices include assessments that

align to objectives and instruction and grading procedures that reflect student

mastery. This study focused on middle school level students' perceptions of

teacher's fairness in assessment-based scenarios.

Fairness and Ethics

The concept of fairness and the concept of ethics have shared tenets, but

the two concepts are not the same. Velasquez, Andre, Shanks, and Meyer

(1996) framed acting ethically as the decision to act morally. They described five

approaches to help a person make an ethical and moral decision. One of these

approaches is the f*airness or justice approach* (Velasquez, Andre, Shanks, &

Meyer, 1996, para. 10). Using the fairness approach, any action is considered

ethical if it "treats everyone in the same way and does not show favoritism or

discrimination" (Velasquez, Andre, Shanks, and Meyer, 1996, para. 10). Other authors agree that fairness is a dimension of ethics. In the *Journal of Academic and Business Ethics*, Stranahan, J. Borg, and Borg (2009) discussed the concept of grading schools using standardized tests. Their paper discussed the possibility of schools' accountability grades being lowered based on the types of students it serves, and argued this was unfair. This concept is consistent with Velasquez, Andre, Shanks, and Meyer's definition of the fairness approach to an ethical decision. Green, Johnson, Kim and Pope (2007) also used the concept of fairness as a dimension of ethics in their paper "Ethics in the classroom practices: Issues and attitudes." The paper discussed the concept of *Do No Harm* (Green, Johnson, Kim & Pope, 2007) as a guiding ethical principle. They presented the idea that teachers not acting fairly may violate this ethical principle:

> Fairness (or protection of student rights) is a general principle that no one contests in the abstract. However, thinking about causing harm focuses the discussion at the level of the implications of everyday practice. Educators must be well versed in the potential impact of the practices they use because their assessment and evaluation may have a variety of unintended consequences for their students. (p. 1001)

Thus, for the remainder of this study, the concept of fairness as a dimension of ethical behavior will be assumed.

Fairness in Education

When a student is at school, his or her teachers are considered to be "in loco parentis," (Popham, 1991). In other words, the teacher acts as the child's parents in the student's best interest. Therefore, it follows that it is a teacher's responsibility to act as a parent by protecting the students' rights. Green, Johnson, Kim and Pope (2007) wrote that "fairness (or the protection of students' rights) as a general principle no one contests in the abstract" (p. 1001). When fairness is considered a right, it should be guaranteed, much as United States' citizens are guaranteed certain rights through the Constitution. Along these lines, Rodabaugh (1996) argued fairness should be institutionalized in classrooms. By doing so, teachers will be guaranteeing the rights of their students, as is their responsibility.

Based on their studies at the university level, Rodabaugh and Kravitz (1994) divided fairness into three categories: interactional fairness, outcome fairness, and procedural fairness. Figure 2.1 presents a definition for each type of fairness and provides examples of each.

The strength of the link between fairness and assessment is reflected in analyzing the three types of fairness. A majority of the examples offered by Rodabaugh and Kravitz (1994) are associated with assessments and evaluation. Procedural fairness almost completely describes assessments, with the exception of attendance, while outcome fairness is entirely focused on assessment and evaluation. Wren, Sparrow, Northcote and Sharp (2009) also found college

students emphasize fairness in describing good assessment and evaluation

practices.

Type of Fairness	Definition	Examples of Fair Behavior
Interactional fairness	The relationship between the students and their professors	• Showing no partiality to some students on the basis of gender, race, or age; • Not being "angry" or "mean" in class (including using profanity, yelling, or screaming); • Not embarrassing students in the classroom, especially by using sarcasm and put-downs; • Exhibiting a caring attitude toward students; • Responding to student questions
Procedural fairness	The rules and regulations that are employed to determine grades, and the policies regarding attendance, make-up tests, cheating and plagiarism, and other student performance	• Rules and regulations related to the classroom to be fair to all students, without exceptions • Tests to be fair and to measure accurately what they have been asked to learn, without trick questions or vague references. • Tests returned promptly and discussed in class, the answers explained, and time allotted for questions. • Monitoring and regulation of cheating and plagiarism.
Outcome fairness	The distribution of grades	• Grades fairly reflect student mastery of the class topics as demonstrated by test scores or the completion of other course requirements. • Individual accomplishments should count toward their final grades, as opposed to group accomplishments

Figure 2.1 Rodabaugh and Kravitz's types of fairness.

In terms of assessment, fairness can be characterized in many ways. Carlson (2003) stated "effective grading practice relies on equitable and appropriate evaluation" (p. 511). So what makes an assessment practice equitable and appropriate? Thinking about an assessment from its inception, the first characteristic would be alignment to instruction. In Wren et al.'s study, they found "there was a high expectation that the assessments set would be closely aligned with the taught elements of the course (2009, p. 14)." The alignment of assessments and instruction is also described as a fair practice by Hill and Zinsmeister (2012) as they described an ethical college professor as being "cognizant of assessments that do not match course objectives" (p. 129).

Second in the cycle of creating an assessment is its design. This may mean avoidance of unfairness in terms of questions being demeaning to any one gender, race, age, or sexual orientation (Rodabaugh, 1996.) Kuhs, Johnson, Agruso, and Monrad (2001) have a broader definition of fairness: "Assessments should be designed to measure the skill(s) of interest and not be affected by the individual differences of students" (p. 5). This includes a person's race, ethnicity, gender, sexual orientation but also includes the background knowledge a student needs to be able to have an equal chance of performing an assessment well. One example Kuhs et al. gave was a question asking about travel to Europe; they proposed this question may not be fair for a low socio-economic child who has never had such an extensive travel experience.

An assessment may also be unfair on the basis of unclear directions, questions, or rubric descriptors (Wren et al., 2009). Sadler (1989) researched the relationship between classroom assessment and teacher expectations. He found students needed to understand the standards used to appraise their work. Thus, for an assessment to be considered fair, the instrument needs to be free from vague wording and easily misinterpreted language. Ideally, the assessment instrument should either be reviewed with the students or made in collaboration with the students. Ross, Rolheiser, and Hogaboam-Gray (2010) found support for this principle when they worked with students to develop a set of performance tasks to appraise students' learning. Then, the instructors and the students both evaluated the students' work after a project using these collaborative assessments. Most students believed that this system "fairly appraised their work" (p.88).

Fairness may also be exhibited in the grading of student work. Holmes and Smith (2003) stated that "students should know from the outset that the grading criteria are the same for all students (p. 319)." This may mean all students, regardless of race, gender, sexual orientation or it may mean regardless of socio-economic status, degree of friendliness, or many other behavioral aspects that have been included in grading by teachers (Marzano, 2000). When a grade or evaluation is changed because of factors that do not show student mastery of the learning objectives (like behavior, attendance,

homework completion, class contributions), the grade is considered polluted (Pope, Green, Johnson & Mitchell, 2009).

False praise may also be considered a form of score pollution. Cameron and Pierce (1994) found students reported greater satisfaction toward a teacher when that teacher had given feedback to the student which was explicitly linked to the quality of the performance, and not to the effort given. In terms of feedback, students in Rodabaugh's study also described prompt feedback as a component to a fair assessment (1996).

Fairness, Trust, and Teacher Effectiveness

As shown in Figure 2.1, Rodabaugh and Kravitz (1994) described one component of fairness as interactional. It is the relationship between the instructor and the students, and whether that relationship is the same for all students. Instructors must be "equally concerned about all students in their classrooms without showing any partiality when answering their questions, when giving them assistance, and when responding through body language or otherwise" (Rodabaugh, 1996, p. 40).

Although many people think of assessment as a written form of communication; however, assessment in the form of feedback can also be given orally or through demeanor. Personal communication can lead to immediate insight about student learning, but needs to have clear learning objectives and judgment criteria identified for it to be effective ("Assessment Methods: A Set of Four Options," 2011).

However, fairness is not perceived solely through the instructor's actions, but also through the assessments he or she designs and uses. In a study of seventh graders, Wang and Holcolme (2010) found that "teachers can best promote students' positive identification with school and stimulate their willingness to participate in their tasks by offering positive and improvement-based praise." Teachers who used assessments as opportunities for constructive feedback maintained a level of engagement with their students which led to student satisfaction of the school environment. However, the converse may be true as well. As Green et al. stated, "teacher–student trust can be damaged by assessments that the student perceives as unfair or unfounded" (2007, p. 1009). A teacher who uses surprise items on a test that did not appear on the study guide is an example from the Green et al. study. They believe this action may do "harm by breaking the implicit bond of trust between teacher and student" (2007, p. 1001). However, there is no research which tests this theory. Can unfair assessment practices diminish the trust a student has in a teacher?

In order for a teacher to be most effective, that teacher must foster a trusting relationship with his or her students. In his studies on skillful teaching, Brookfield (1990) found a "teacher's actions, and the trust these inspire or destroy, are crucial to learning" (p. 163). Educators must be well versed in the potential impact of the practices they use because their assessment and evaluation may have a variety of unintended consequences for their students (Black & Wiliam, 1998).

Researchers have found middle school students who perceived their teachers "cared" were more likely to be motivated to succeed (Wentzel, 1998). These same caring teachers were identified as those who acted in what Rodabaugh would call a fair way: they showed interactional fairness through democratic principles, procedural fairness through developing expectations for student behavior despite individual differences, and outcome fairness by providing constructive feedback (Wentzel, 1997).

In terms of student demographics and trust, many studies have found females tend to have more of a relationship with their teachers compared to males. In contrast, males typically have more conflict and less closeness in their relationships with teachers than females (Baker, 2006; Howes et al., 2000; Hughes, Cavell, & Wilson, 2001). However, these studies were all based on students in pre-school through sixth grade. Therefore, this study will attempt to provide or more complete picture of the gender and trust relationship by using the student's gender as a factor during analysis.

Student Perception of Assessment Techniques

In designing classroom environment that is fair, including the evaluation procedures and assessment measures, the teacher's perspective may differ from the students' perspective. The teacher may believe he or she has created an interactional, procedural, and outcome fair curriculum, but his or her students may not share that perception. Brookhart and DeVoge (1999) described the classroom assessment environment as a shared concept: it is created by not only

the actions of the teacher but also the meaning the students derive from those actions.

Many studies show the difference between student perceptions and teacher perceptions of a shared reality. In Gulikers, Bastiaens, Kirschner and Kester (2014), the researchers discovered both freshmen and senior high school students had a different idea of the degree of authenticity in an assessment task than their teachers. When looking at the same assessment tasks, teachers believed the task, method, and criteria of the individual assessments were more authentic than the students did. The students' perceptions were similar, but different from the teachers' view.

Therefore, investigation of student perceptions on fairness would be a logical next step. In this line of research, Gordon and Fay (2010) examined several hundred college students' perceptions of fairness in their evaluative systems. They found that college students believed interactional fairness (the instruction given in class) to be more important than outcome fairness (how teachers used assessment to create the course grade).

In another study of student perceptions, Alkharusi (2010) examined ninth grade student self-efficacy. He examined the perceived assessment environment of these high school freshmen, one in which assessments were challenging, clear, and provided detailed feedback versus classrooms in which assessments were difficult, feedback was in the form of a grade, and assessments were used to highlight social comparison. Alkharusi (2008) called these environments

20

learning-oriented and harsh-oriented, respectively. However, at the conclusion

of his study, he found a limitation: it made an omnibus categorization of

classroom as either learning-oriented or harsh-oriented. He continued in saying,

this might have made it difficult to know which aspects of the recommended

"classroom assessment practices could be considered responsible for the

observed effects detected in the study" (p. 38).

Assessment Practice Literature

In order to measure students' perceptions of fairness, the students need

to have plausible scenarios on which to base their opinions. In development of

surveys, an initial step is to investigate the literature on best practice in

assessment in order to create scenarios which are both aligned with Rodabaugh's

types of fairness and reflect best assessment practices. The following sections

describe the results of that best practice investigation.

Grading formative assessment.

Formative assessment is described by Stiggins (2008) as assessment for

learning. In other words, teachers use formative assessment to diagnose their

day to day instruction, both to gather data on how effective the lesson was, but

also to analyze which students might need further help. Formative assessment is

a time when students should be allowed to make mistakes, so the teacher might

discover misconceptions early and address them. Referring to formative

assessments, Stiggins stated, "These have no place in the gradebook," (2008, p.

35). Green and Johnson also agreed formative assessments should not be

graded, but elaborate as to why this concept is difficult for teachers to accept: "Many teachers…believe that students will not take the work seriously unless it is graded" (2010, p. 111). Thus, student achievement is often measured using a grade before any substantive feedback is given to the students and before any teacher guided practice is provided.

Using student achievement measures to change behaviors.

 One of the goals of the educational system is to develop students who will contribute to a democratic society in a positive way as they become adults (Green & Johnson, 2010). Thus, many teachers argue that they must teach non-academic factors such as participation, following directions, punctuality of assignments, effort, attendance, work habits and neatness, and attitude (Guskey, 2009). Assessment experts agree that these factors are important to educating the whole child, but feel they could assign these as separate grades, what Guskey calls "process grades" (2009, p. 21).

However, many teachers are hesitant to separate those factors because they are used to change behavior, either as a reward or a punishment. "Teachers perceive non-achievement factors such as effort, behavior and attendance as important to classroom control, and consequently, often include them in their grading practices" (Marzano, 2000, p. 37). For example, McMillan, Myran and Workman (2002) provided an example where rewarding effort might lead students "who are not competent…to believe that they demonstrated needed knowledge or skills" (p. 212). Without the effort grade, the teacher

might think the child might not try, which makes learning more difficult. On the punishment side, many teachers take away points for work being turned in late. Thus, a student who might completely understand the concepts, but is disorganized, might appear to not have mastered the learning objectives. These type of measurement practices affect validity. As Green, Johnson, Kim and Pope (2007) described the problem: "the score on the test does not represent actual student achievement in the content area and is 'polluted' by factors unrelated to academic attainment" (p. 1001).

Using zeros in measuring student achievement.

In mathematics, a zero represents the concept of nothing. When we consider student achievement after appropriate instruction, it is a rare instance for a student to have learned nothing (Raebeck, 1993). Most often, zeros are used as punishment for what the teacher perceives as a lack of responsibility (Guskey, 2009) because it reflects missing work. It is mathematically correct: there is no work, so it is a zero. However, in terms of that zero representing student achievement, it rarely leads to a valid conclusion.

Also in terms of mathematics, a zero has an "undeserved and devastating influence" (Guskey, 2009, p. 137) on a student's grade. A zero creates a much larger interval between a D and an F, compared to the other grade intervals. For example, in South Carolina, the distance between a zero and a D is 70 points, whereas the distance from a D to a C (and all subsequent sequential letter grades) is only 7 points. As Guskey (2009) said, the use of a zero in grading

"mathematically and ethically is unacceptable" (p. 138). Given that strong wording, the current study will include inquiry into student perceptions on the use of zeros.

Using any type of assessment exclusively.

As students progress to middle school, their instruction turns from an integrated model (where they typically had one or two teachers to deliver the four major subjects) to a content specific model (where they have four teachers each of whom specializes in one of the four major subjects.) This shift causes teachers to become content experts. As content experts, middle school teachers also tend to assess students in more rigid ways that conform to their content. For example, math is a subject that is easily assessed in many ways through selected response assessments, such as multiple choice or true/false questions. Writing, however, is most easily assessed through performance tasks. However, as Gronlund (2006) discussed, "effective assessment requires that a variety of assessment procedures be used" (p. 19). By using a myriad of assessment types, a teacher is increasing the validity of his/her instructional decisions. Wormeli described this idea: "When we assess students through more than one format, we see different sides to their understanding" (2006, p. 31). Thus, a survey on assessment and fair practices should include questions about the use of a variety of assessment types.

Evaluating non-content area standards with content standards.

A grade is a means to measuring "how well a set of instructional objectives have been achieved" (Gronlund, 2006, p. 169). In the instructional process, many advocates of assessment recommend a backwards model, in which teachers identify the learning outcomes (i.e., objectives), decide how those assessments can be most effectively assessed, and then create the instruction that will help guide the students to success on the assessment (Wiggins & McTighe, 1998). Thus, the assessments and the learning objectives are closely aligned. However, this ideal situation does not always occur. At times, teachers will create or use assessments which are meant to measure one objective, but require mastery of a different content area. For example, a student might understand and be able to show mastery on the aspects of the French Revolution, but cannot read the test. Thus, the child receives a poor grade on the social studies test, but that poor grade truly indicates a deficit in reading, not social studies. Similarly, one content area assessment might measure its content objectives, but also include non-content objectives that seem relevant. One instance might be a science project which requires writing. The teacher might have a rubric which is measuring the various science concepts, but that same rubric might also include non-science content such as handwriting, spelling and punctuation. When this occurs, the science grade becomes a writing assessment. Assessments should illuminate, not obfuscate, the achievement a student has made toward mastery of a learning objective

(Wormeli, 2006). This a scenario that depicts the muddling of content areas will be included on the survey.

Developing the Survey Instrument

In this study, a Likert question followed each scenario to measure the students' perceptions of fairness. The researcher decided to use six-point response categories for several reasons. First, the researcher wanted to use an even number of responses to avoid a neutral response. Odd number responses allow for a middle point, which might prevent this researcher from discriminating between those students who perceived a scenario as fair or unfair and those scenarios which build or undermine trust (Johnson & Morgan, in press). Also, six-response categories were used to increase reliability and validity. In their study on the optimal number of response categories, Preston and Colman (2000) found reliability as measured by Cronbach alpha coefficients increased as the number of responses increased; but after seven responses, the coefficients did not increase at a statistically significant level. They also reported that validity and ability to discriminate between respondent choice increased as the number of response categories increased. Thus, six-point response categories were used as this format supports validity and reliability with an even number of possible responses.

Additional consideration was taken in preparing the response scales. First, response categories were ordered from negative responses to positive responses based on the writings of Fink (2003).   Other research supports Fink's

stance. For example, Nicholls, Orr, Okubo and Loftus (2006) found respondents who completed a questionnaire that started with the statement 'definitely agree,' responded with that answer 27 percent more than did the students who were offered 'definitely disagree' as their first option. The scale also had what Johnson and Morgan (in press) call opposite end points, with answer choices that "reflect equal intervals along a continuum" (p. 18). The remaining four points were mirror-images of each other, with a total of three positive responses and three negative responses.

Purpose of the Study

Fairness as a construct has been described in the academic literature. However, there are three aspects to fairness which have not been explored. First, quantitative measurements concerning which assessment practices are considered fair by students are sparse. Second, in the few assessment studies that address fairness, the students are often at the university level. Finally, no empirical evidence appears to exist on the effect of fairness, or lack of fairness, in relationship to student-teacher trust. Therefore, this study examined, in a quantitative manner, how middle school students perceive various assessment scenarios in terms of fairness. These results were also used to investigate the relationship between a student's perception of fairness and the change in trust that child has in his/her teacher based on the assessment scenario.

- The first hypothesis for this study was that there will be a difference in fairness perception scores, based on the underlying best practice

27

characteristic of the scenario, but not based on the type of fairness (i.e.,

interactional, procedural, and outcome) represented in the scenario.

There may indeed be differences in students' responses based on type of

fairness, because the types of fairness appear to represent different aspects of

teaching, assessing, and evaluating. Interactional fairness depicts more of the

formative assessment, day to day evaluations, of student performance.

Procedural fairness outlines the summative assessment practices of a teacher,

the ways in which a teacher assesses a unit of study. Outcome fairness deals

with how a teacher gives a final grade, a complete evaluation, for each student.

These are different aspects of teaching and assessing, so the students'

responses may show they are more sensitive to a teacher acting, in their

perception, more fairly during the formative, summative or evaluative period of

instruction. However, for the purposes of this study, while those differences will

be explored and discussed, the null hypothesis states the students' responses will

vary only by the underlying fairness characteristic (i.e. deviates from best

practice or promotes best practice) rather than the type of fairness.

- The second hypothesis is the students' perceptions of fairness will have a
  positive relationship to the students' change in trust perception.

Chapter Three

Method

The following sections describe the methods that were used to investigate students' perceptions of fairness, and to study any relationship between the students' perception of fairness and the students' perceptions of trust. Explained below are the measures, samples for focus groups and surveys, and data analysis and procedures that were used in this study.

Measures

This study measured two aspects of student perceptions. First, it examined students' judgments of various assessment scenarios in terms of fairness. Then, it investigated if there is a relationship between the students' perceptions of fairness and the trust the assessment practice in the scenario would create. For example, a scenario may have described a teacher who grades based on the length of their essays, and not on the content. The first question students answered asks about their perception of the fairness of this practice; the second inquired as to how this practice would influence their level of trust in the teacher.

To ensure a balanced presentation of fairness issues, the researcher first created a table of specifications for the survey (Table 3.1). Like a traditional table of test specifications, this step ensured the survey questions inquired into

29

different types of scenarios and did not include "superfluous assessment items" (Green & Johnson, 2010, p. 56). There were an equal number of scenarios for each type of fairness (i.e., interactional, procedural, outcome). There was one more scenario which deviates from best practice than promotes best practice.

Table 3.1

Table of Test Specifications for the Survey

Alignment to Best Practices	Interactional Fairness	Procedural Fairness	Outcome Fairness
Promotes	9	6	3,5
Deviates	1, 2	4,7	8

Note. Number represents the scenario number.

The survey was created based on the table of test specifications for the survey. It consisted of an introduction and short description of the study, followed by one scenario. After the scenario, two questions were asked: "If you were a student in this class, how fair is this scenario?" and "If you were a student in this class, and this actually happened, how would it change your trust of the teacher?" Both questions were followed by six-point Likert responses, with lower numbers representing the extreme "unfair" or "damage trust" categories, respectively. An example is provided below, in Figure 3.1, and the entire survey and scenarios can be reviewed in Appendix B.

To increase the validity of the survey, nine educators (i.e., middle school teachers and administrators) provided feedback on the scenarios. Their responses provided evidence based on response processes, which helped the researcher discover if there was a "fit between the construct and the detailed

nature of the performance or response actually engaged in by test takers"

(American Educational Research Association, p. 15). They were asked to

comment specifically, for each scenario, "Is this clear to a typical middle

schooler? Is it possible this might happen in a middle school classroom? Can I

make it better? If so, how?" The researcher used the educators' responses to

clarify the scenarios before the survey was sent to the students.

Sample Scenario and Questions

One of your teachers is trying to motivate everyone to do their best. One way
he is trying to increase the students' motivation is by handing back tests to
students in order of how well they did. Students with the highest grades get
their tests back first, and students with the lowest grades get their tests back
last.

How fair is it for the teacher to hand back tests in order from highest to
lowest?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your
teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Figure 3.1. Sample survey scenario and questions

Samples

This study was conducted in a suburban school district in the Southeastern United States with approximately 27,000 students in pre-school through twelfth grade. The district has 18 elementary schools, 7 middle schools and 5 high schools.

This study examined the perceptions of fairness in assessment scenarios and the relationship between these perceptions and change in the student-teacher trust relationship. The middle grades were selected for this study for two main reasons. First, most studies addressing fairness and assessment focused on university students. Thus, in this study the researcher wanted to explore the perceptions of students in K-12. Also, in middle school, students become adolescents. As Eccles and Midgley (1989) found, teacher and student relationships decline as students make the transition into adolescence. Thus, this study was designed to provide information that might inform the role of assessment in contributing to that declining relationship. The sample was further restricted to seventh grade students; at the time of the study, they had completed one half of their middle school experience, which signifies they are equally removed from elementary and high school.

The district's Research, Accountability and Evaluation department provided student demographic data, as well as school rosters. It also served as the survey administrator, both sending out the survey and receiving the results. Recent enrollment data showed there were 2,150 seventh graders in the

participating district. Of those students, 53 percent were male, 47 percent were female. Three percent of students were Asian, 60 percent were African-American, 7 percent were Hispanic, 27 percent were White, and the remaining three percent were either American Indian, Native Hawaiian or identified themselves as two or more races.   Forty nine percent of the students received free or reduced lunch.

In order to help ensure the scenarios and questions of the survey would be understood by the students, the researcher first conducted a pilot test. Fink (2006) suggested "trying out the survey process initially by using a different method from the one you eventually intend to use" (p. 40). In this case, the survey instrument was first administered to a focus group in interview form. In order to capture "a sample that represents various groups and patterns of characteristics in the desired proportions" (Fink, 2006, p. 49), a focus group was conducted which was comprised of seventh graders from a middle school in the district whose seventh graders most matched the district's seventh grade population in terms of gender, race and socioeconomic proportions, which was School C in Table 3.2. The table shows those demographic data points. Once the focus group from School C gave feedback, the instrument was revised. Changes made to the instrument are discussed in more detail in Chapter Four. Subsequently, the remaining 2,150 seventh graders in the district were given the revised survey.

Table 3.2

Seventh Grade Demographic Data in Percent from District's Middle Schools

Location	Male	Asian	African-American	Hispanic	White	Free/Reduced Lunch Status
District	53	3	60	7	27	49
School A	53	6	57	9	23	57
School B	53	4	66	9	18	56
School C	50	3	59	6	27	48
School D	57	3	45	3	46	26
School E	48	4	64	7	23	48
School F	55	0	80	5	11	62
School G	54	0	49	5	43	38

Data Analysis

In order to support validity, the survey scenarios were first presented to nine educators with professional middle school experience as well as a focus group comprised of seventh graders. This step was necessary in providing content validity evidence, or as Doherty (2008) stated, when "you want to improve something, ask the people who do it" (p. 82). One focus group was used, with seven randomly selected seventh grade students. Brown (1999) says that the group should consist of 4-12 if the group is homogeneous and 6-12 if heterogeneous; a balance between the need to have enough people for a lively discussion and the danger of an overwhelming group size must be achieved. The selected students' parents/guardian received an invitation letter, describing the study, explaining the benefits and risks, and asking for permission. Because the students are minors, receiving adult permission was necessary legally and ethically. However, the researcher also asked for the child's permission on the

34

same letter. Once the students had been selected and permission had been given, the researcher coordinated with the school to meet with the students during non-instructional time, in this case, lunch. The researcher used the script and questions given in Appendix A. By asking for feedback before the large-scale data collection, the researcher was able to refine the questions. As Glesne stated, this process "enables [the researcher] to focus and shape the study as it proceeds" (2006, p. 148). The researcher took notes during the focus group. Later, the researcher used coding to "break down, examine, compare, conceptualize and categorize" (Strauss & Corbin, 1990, p. 61) the data from the interviews. These patterns then were used to change the scenarios in order to provide clarity or relevance. The data gathered helped the "choice of question wording become more objective and less a matter of research judgment" (Fowler, 2002, p. 6).

The first research question of this study investigated student perceptions of fairness based on scenarios that were constructed according to best practice in assessment and to exemplify one of the three types of fairness as described by Rodabaugh (1996).

1. Do students' perceptions of fairness in an assessment scenario differ based on the best practice characteristics of the scenario (if it is written to promote best practice or deviate from best practice), but not based on the type of fairness (i.e., interactional, procedural, outcome) represented?

To answer this question, first a fairness alignment coefficient (FAC Sub-score) for each student was created, in regards to type of fairness. The FAC sub-score was a sum of the student's perception of fairness for each of the three scenarios that represented the three different types of fairness. In other words, the students' perception of fairness for scenarios 1, 2, and 9 were added together to create a FAC-interactional score. Similarly, scenarios 4, 6, and 7 were added together to create a FAC-procedural score and scenarios 3, 5 and 8 were added together to create a FAC-outcome score. Each student had three FAC sub-scores, and one total FAC score.

However, because some scenarios were intentionally written to deviate from best practice, the student's responses for these items were reverse coded. Thus, after this recoding, a score of six represented a student either (a) perceiving a scenario deliberately written to promote best practice as completely fair or (b) a student perceiving a scenario deliberately written as deviating from best practice as completely unfair.

Then, to determine if the FAC for each sub-score is similar, a within-subjects analysis of variances (ANOVA) was performed. A within-subjects' analysis is necessary because each student's perceptions of fairness was represented in multiple types of fairness' alignment coefficients sub-scores. An ANOVA was desirable because there were three types of fairness sub-scores represented in the scenarios. The dependent variable was comprised of the FAC sub-score for each type of fairness. The three levels of the within-subject factor

36

are interactional, procedural, and outcome fairness. When significant differences were found ($p<.05$), Tukey HSD post hoc procedures were applied as necessary to determine where the differences in FAC sub-scores lie.

While the scenarios were developed specifically with Rodabaugh's types of fairness in mind, to provide confirmatory validity, factor analysis was completed on the results. Ideally, the students' responses to the interactional, procedural and outcome scenarios would have had similar patterns of responses because they are all associated with a latent variable-type of fairness. However, the researcher measured the relationships between the scenarios and reported the results of the factor analysis.

The second research question of this study investigated if there was a relationship between students' perception of fairness, student's gender, and change in trust student perception.

2. To what extent does a student's perception of fairness and the student's gender predict his or her perception of change in trust toward the teacher?

Each student's reverse coded Total Fairness Alignment Coefficient was used as one predictor. The student's gender was used as the second predictive variable. The trust variable was reverse coded for scenarios which were intentionally written to deviate from best practice. Thus, after this recoding, a score of six represented (a) a student gaining trust based on a fair practice or (b) a student losing trust based on an unfair practice.

This regression coefficient addressed the second hypothesis that students whose perceptions of fairness were more in line with best practice also changed the level of trust for the teacher in a significant manner. In other words, a student who found scenarios that deviated from best practice as more unfair, and who found scenarios that promoted best practice as more fair, also changed the trust in the teacher in a similar pattern (less trust for unfair practices and more trust for fair practices). Also, by using gender as another predictive value, the model tested the question "Did gender have a role in these changes of trust?"

Procedures

For the focus group, seven students were randomly selected from School C's seventh grade population. The researcher met with these students during their lunch period and used the script and questions listed in Appendix A. The purpose of the focus group was, as Fowler states, "to compare the reality about which respondents will be answering questions with the abstract concepts embedded in the study objectives" (2002, p. 106).

Using data from the focus groups, scenarios were honed for the larger survey. Then, all seventh grade students in the district received the survey via the K12 Insight survey software program used by the district. Students were told the survey was not mandatory and their responses would be kept confidential. The survey began with an introduction, and description of the survey. Students were given twelve days to complete the survey, which was

sent via their school email, and was tracked using the K12 software's program so students could not complete it more than once. All data were kept in a password-secured computer and network throughout the collection and analyses.

The data were divided into two tables. The first identified the individual student's responses concerning fairness to all nine scenarios. It recorded the student's unique identification number, the scenario number, and it noted if the situation was written to promote best practice, or to deviate from best practice. These pieces of data were useful first in descriptive statistics. Such statistics as the overall mean scores for each scenario were reported, as well as overall fairness scores for scenarios that were written in concordance or discordance from best practice. The data table also indicated which type of fairness the scenario represented. These data were used to create the FAC sub-scores, which were then used in the within-subjects ANOVA.

The second data table was similar to the first data table, in that it also recorded the student's unique identification number, the scenario number, and it noted if the situation was written to promote best practice, or to deviate from best practice. However, it also recorded the change on trust perception as well. These data were first used to create descriptive statistics such as the mean in student perceptions of fairness and mean in student perceptions of change in trust for each scenario. However, it was also used to run a multiple regression analysis with student perception of fairness and student gender as predictive

variables for the change in trust variable. Both analyses were created using

Statistical Package for the Social Sciences (SPSS), version 21.

Chapter Four

Data Analyses and Major Findings

This chapter presents a variety of data analyses designed to answer the research questions:

1. Do students' perceptions of fairness in an assessment scenario differ based on the best practice characteristics of the scenario (if it written to promote best practice or deviate from best practice), but not based on the type of fairness represented (i.e., interactional, procedural, outcome)?

2. To what extent does a student's perception of fairness and gender predict his or her perception of change in trust toward the teacher?

Qualitative results from the educator surveys and from the student focus group will be reported first. The descriptive and inferential statistics concerning student perceptions of fairness will follow. Finally, analyses of students' perception of trust, student gender, and students' perception of trust will conclude the chapter.

Qualitative Results from Surveys and Focus Groups

In order to increase evidence of validity based on response processes, the researcher gathered input from stakeholders familiar with the types of behaviors portrayed in the scenarios. Seven seventh grade students from a school that is

similar in district demographics participated in a focus group. Specific

demographic data describing the backgrounds of these students can be found in

Table 4.1. The researcher followed the protocol as described in Appendix A.

Table 4.1

Focus Group Demographics

	Number of Students
Gender	
Female	3
Male	4
Lunch Status	
Free or Reduced	4
Full Pay	3
Race	
African-American	4
Latino	1
White	2
Reading Achievement	
1-25th percentile	1
26th-50th percentile	2
51st to 75th percentile	1
75th to 99th percentile	3

Note. Lunch status is an indicator of socio-economic status, and represents is a child's parent/guardian has applied for and has been approved for free or reduced lunch based on family size and income. Reading achievement was measured using the Measures of Academic Progress reading assessment. The most recent score was used.

Nine educators who either are current middle school teachers or

administrators were also sent a copy of the scenarios, and were asked the same

questions from the focus group. Their responses were anonymous in hopes of

more honest responses.

Generally, the stakeholders felt the scenarios were both clear and

conceivable based on the stakeholders' past experiences, which were the goals.

Overall, seven of the scenarios were deemed completely clear and plausible by the educators. They had some wording suggestions, for example, changing "reading teacher" to "ELA teacher" in scenario five because middle schools do not have reading teachers. However, scenarios one and two caused mixed responses from the educators. Five out of nine educators believed a middle school teacher would (and some indicated has) marked down a student's grade significantly based on handwriting. Four of the nine could not conceive of such an event occurring. One made a suggestion of changing the scenario from down grading the work from an A to B to down grading it from a 100 (highest A) to a 93 (lowest A). This suggested change was made and presented to the student focus group. Interestingly, all of the student focus group members believed scenario one to be completely plausible, and had, in fact, happened to either themselves or someone they knew. Their suggestion, which was made without any prompting from the researcher or knowledge of the first draft of the scenario, was to have the teacher change the grade from an A to a B, which was how the scenario was originally written. The researcher decided to follow the students' suggestions, as opposed to the educators, since every student in the focus group was in agreement, and because the population being studied was the students.

The other scenario which the educators could not agree upon was scenario two, which describes a teacher handing out papers from top scores to bottom scores. Seven of the nine educators viewed the practice as plausible,

while two could not imagine it happening. This question also caused more comments than any other. One comment which was representative of many stated, "I hate this system.....I know some teachers do this...I think it is a terrible way to motivate kids!" (Anonymous, 2015). Again, when the researcher asked the student focus group about scenario two, each child found it plausible, which led the researcher to leave the scenario in the survey, unchanged.

The students, including the student at the 8th reading percentile, were all in agreement for each scenario. They believed each was clear and plausible. The researcher was especially interested if the student at the 8th reading percentile could understand the text of the scenarios, because there are many students reading significantly below grade level in the seventh grade throughout the district. The student was chosen to restate or describe three of the nine scenarios, and could do so well each time. This helped bolster the researcher's confidence that all students would be able to read and understand the scenarios.

There were wording suggestions they felt would improve the scenarios. They agreed "nicer" was not the best word to use in scenario five, as "nice" was too vague a word. They suggested it be changed to "cooperative." The researcher agreed and made the change before the final survey went out. The other change the students suggested was changing the French teacher to a mathematics teacher. They informed the researcher that the practicing of oral word formation is not common in their level of foreign language; that they are mainly memorizing vocabulary words. When asked if the students could see the

point of the scenario, they described it as "the teacher wants the students to make mistakes during class, but not on the test" (Anonymous, 2015), which was exactly the purpose of the scenario. The students then suggested the teacher be a math teacher, since they often have opportunities to try a new math concept in class before seeing it on a test. The researcher agreed, and, while keeping the frame of the scenario intact, changed details to reflect a math classroom.

Survey Results

After the focus group's and educator survey's results were reviewed, changes as described in the previous section were made. The survey was then sent to 2,141 seventh graders currently in the district. A response rate of 34.5% was attained, which was 738 responses.

Psychometric Qualities for Fairness Questions

Table 4.2 shows the students' perceptions of fairness for each scenario. The complete scenario can be found in Appendix B. For five of the nine scenarios (numbers 2, 3, 4, 6, and 9), over half of the students reported perceptions of fairness which aligned to best practice (either they found practices which deviated from best practice to be unfair, or practices which promote best practice to be fair.) One scenario, number 8, which describes a teacher who only uses one type of assessment, found students evenly divided between fair or unfair. Finally, for three of the nine scenarios (numbers 1, 5, and 7), over half of the students reported perceptions of fairness which did not align to best practice

(either they found practices which deviated from best practice to be fair, or

practices which promote best practice to be unfair.)

Table 4.2

Student Perceptions of Fairness Results Shown as a Percentage

Scenario Number	Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair
1	8.51%	14.09%	14.78%	7.81%	10.18%	44.63%
2	27.03%	18.51%	20.41%	11.08%	13.38%	8.78%
3	10.54%	9.46%	8.92%	9.46%	21.76%	39.05%
4	21.40%	19.34%	17.01%	11.52%	12.21%	18.52%
5	17.28%	14.68%	20.30%	9.47%	14.95%	23.32%
6	7.25%	7.80%	9.30%	9.85%	15.87%	49.93%
7	13.16%	13.43%	12.19%	11.22%	14.68%	35.32%
8	11.42%	16.64%	22.00%	13.68%	17.77%	18.48%
9	12.10%	12.94%	16.60%	12.38%	15.89%	30.10%

These percentages represent the students' responses to the survey

questions concerning fairness. However, in order to complete inferential data

analysis, the results first had to be coded as continuous variables, as designated

on the survey (6 representing "Completely fair" and 1 representing "Completely

unfair," for example). Then, the scores of any questions which deviated from

best practice needed to be reverse coded. This change supports better clarity in

analyses, as a "6" would now represent a student's perception toward fairness as

completely aligned with the scenario's promotion or deviation from established

best practice.

After recoding the appropriate questions, the researcher examined the

alignment to best practices of each scenario, known as the Fairness Alignment

Coefficient, or FAC. Table 4.3 shows these results. Four of the nine scenarios

showed students' perceptions of fairness (FAC's) were generally aligned to best practice as their overall means were greater than 4.0. Scenario six, which allowed the students to use calculators on a Social Studies timeline assignment, had the greatest alignment, with $\overline{x}$=4.69. Three scenarios showed some alignment to best practice as their overall mean scores were between 3.5 and 3.9.

However, there were two scenarios in which the students' perceptions of fairness (FAC) did not align to best practice. First, scenario seven, which described a student who is an avid reader but is unorganized with her reading log, had the lowest FAC ($\overline{x}$=2.93). Over 35% of the students felt the teacher would be acting completely fairly in giving the girl in the scenario a zero for not turning in her reading log. However, research into best assessment practice scenarios warns against such teacher actions as the grade does not represent the child's achievement in the content (in this case, reading); instead, it represents the child's unwanted behavior of forgetfulness and disorganization.

The other scenario which showed a lack of alignment was scenario eight. In this scenario, a teacher gives only projects to assess her students. When given the idea that there is a student who prefers selective response assessments, the question was asked if it is fair for the teacher to only give open-ended projects even for a child whose learning style performs better with a more defined assessment type. Students were in slightly out of alignment with best practice, $\overline{x}$=3.35.

Table 4.3

Student Perceptions of Fairness for Each Scenario (FAC)

Scenario Number	Type of Fairness	Best Practice	Mean Student Perception	Standard Deviation for Perception
1	I	D	4.50	1.71
2	I	D	4.08	1.65
3	O	P	4.41	1.75
4	P	D	3.71	1.80
5	O	P	3.60	1.81
6	P	P	4.69	1.66
7	P	D	2.93	1.85
8	O	D	3.35	1.65
9	I	P	3.97	1.77

Note. In the Type of Fairness column, P stands for procedural, I stands for interactional, and O stands for outcome. In the Best Practice Column, D stands for deviates and P stands for promotes.

To examine internal consistency, Cronbach's alpha was calculated for the fairness questions ($\propto$ = 0.45), which is considered to show low reliability. Discrimination indices were calculated for survey questions as well. Table 4.4 presents these indices.

Table 4.4

Discrimination Indices for Fairness Items on Assessment Scenarios Survey

Scenario Number	Discrimination Index	Cronbach's Alpha if Item Deleted
1	0.33	0.37
2	0.20	0.46
3	0.22	0.34
4	0.18	0.47
5	0.19	0.38
6	0.23	0.34
7	0.26	0.42
8	0.32	0.38
9	0.26	0.30

Most of these indices (seven out of nine) indicate these items discriminated somewhat between students with high FAC's (i.e. high scores when discussing a scenario which promotes best practice and low scores when describing a deviation from best practice) and students with low FAC's (i.e. scores that were lower when discussing a scenario which promotes best practice and higher when a scenario which deviates from best practice was presented). The first item had a discrimination index of 0.33, which indicates that students who believed the handwriting item to be more unfair were more likely to have higher FAC's overall. Similarly, the eighth item had an index of 0.32 indicates that students who believed assigning only projects to be unfair were more likely to have overall FAC scores. The Cronbach's alpha would decrease if seven of the nine scenarios were removed, and the Cronbach's alpha would only increase by 0.01 if scenario two was removed and by 0.02 if scenario four was removed.

The Cronbach's alpha of 0.45 might indicate students' perceptions of fairness were not merely based on the promotion or deviation of the scenario from best practice. If it were, the internal consistency should be much higher and would represent one common latent factor: alignment to best practices literature. Therefore, as part of an examination of validity, an exploratory factor analysis was performed to investigate what the latent variables might be, since it does not seem solely to be alignment to literature. Using the results from the fairness questions, Kaiser-Meyer-Olkin Measure greater than 0.6 (0.68) indicates the sample is adequate for exploratory factor analysis. The Bartlett's test of

sphericity also surpasses the adequacy test, as it is significant at the <.0005 level. An orthogonal rotation was used as the researcher assumed the latent variables were uncorrelated. Additionally, coefficients that were smaller than 0.32 were suppressed.

The exploratory factor analysis revealed three latent concepts in the student overall Fairness Alignment Coefficients. The scenarios which loaded together on factor one were numbers 1, 5, 7, 8, 9. These scenarios described marking down due to poor handwriting; allowing a student to earn a grade based on work, not attitude; giving a student a zero for missing work; a teacher who gives only projects; and a teacher who only grades summative assessments.

Scenarios 3, 5, 6, and 9 were loaded together on factor two. These describe a teacher refusing to give a zero for missing work; allowing a student to earn a grade based on work, not attitude; giving students who struggle with math calculators for a social studies assignment; and grading only summative assessments.

Finally, factor three had scenarios 1, 2, and 4 loaded upon it. These involve marking down due to poor handwriting; handing back results from high to low; and a teacher grading homework as a means to encourage students to do homework.

Looking at these results, factor one may represent a latent variable of justice. Fairness and justice stem from the same family, but are not the same concept. The scenarios in factor one all seem to share the concept of what a

person deserves. For example, in scenario one, does the child deserve a lower grade due to a trait that is not under his or her control? In scenario seven, does the girl who has had many warnings about her late reading log deserve a zero when she is late one more time? This construct seems to be version of middle school justice.

Factor two seems to have a latent-construct of empathy. These four scenarios all describe students who struggle with something: organization, social skills, subtraction, learning a new skill.   By loading onto the same factor, it seems students perceived these scenarios in a similar manner, indicating they may feel when a teacher is acting empathetically, the students will feel the actions are fairer.

Scenarios 1, 2, and 4 load together to make factor three. These three scenarios seem to have a latent construct of control. In each, the teacher is attempting to control the student by means of a grade.   The child must control her handwriting or risk a lower grade, do better on the test to not have it handed back last, or do the homework unless he or she would like a zero in the grade book. Handwriting, studying or effort, and time management at home are all outside of a teacher's control. By loading these three scenarios onto one factor, it appears student responses similar when a teacher attempts to regulate student actions.

Table 4.5

Exploratory Factor Analysis Structure Matrix

Scenario Number	Type of Fairness	Best Practice	Factor 1	Factor 2	Factor 3
7	P	D	.72		
8	O	D	.66		
9	I	P	-.63	.35	
1	I	D	.61		.34
6	P	P		.73	
3	O	P		.69	
5	O	P	-.35	.39	
4	P	D			.77
2	I	D			.74

Note. In the Type of Fairness column, P stands for procedural, I stands for interactional, and O stands for outcome. In the Best Practice Column, D stands for deviates and P stands for promotes. Cells that are empty indicate the scenario does not align to the factor at a value higher than 0.32.

Hypothesis Testing: Fairness Perceptions

In order to determine if Rodabaugh's types of fairness influenced the results, a within-subjects analysis of variance was performed. First, the researcher checked for assumption violation by conducting a Mauchley's test for sphericity (which concluded that the data do not violate the sphericity assumption with a value of p=.34), a Shapiro Wilk test for normaility (which concluded that the data come from a normal distribution with a value of p=.32), and a test for outliers. Once it was determined there were no violations of assumptions, a within-subjects ANOVA determined that mean student perception coefficients differed at a statistically significant level between types of fairness $(F(2,1314)=48.36, P<.0005)$. Post-hoc tests using the Bonferroni correction revealed that the mean value for students' perceptions of fairness involving

scenarios representing procedural fairness (FAC-Procedural) was slightly less than students' perceptions of fairness involving scenarios representing outcome fairness (FAC-Outcome) (11.30 ± .12 vs. 11.41 ± .11), which was not statistically significant (p=1.0). However, the mean value for students' FAC-Procedural was statistically significantly less (p<.0005) than students' perceptions of fairness involving scenarios representing interactional fairness (FAC-Interactional) (11.30 ± .12 vs. 12.61 ± .11). Also, the mean value for students' FAC-Outcome was statistically significantly less (p<.0005) than students' FAC-Interactional (11.41 ± .11 vs. 12.61 ± .11). Thus, it does seem students do seem to recognize fair or unfair situations (as defined by best practice) when they are presented in an interactional Scenario, one which involves the relationship between the students and their teachers, more so than when the fair or unfair action is presented in a procedural or outcome scenario.

Psychometric Qualities for Trust Questions

The second research question examines if the students' perception of fairness and/or the student's gender can predict the trust a student will feel after experiencing the different assessment scenarios. To examine internal consistency, Cronbach's alpha was calculated for the trust questions ($\propto$ = 0.57), which indicates a low level of reliability. Discrimination indices were calculated for survey questions as well. Table 4.6 presents these indices.

All but one item revealed good discrimination (>.30), which indicates that for each of these scenarios, students whose trust increased or decreased,

Table 4.6

Discrimination Indices for Trust Items on Assessment Scenarios Survey

Scenario Number	Discrimination Index	Cronbach's Alpha if Item Deleted
1	.39	.54
2	.33	.57
3	.34	.44
4	.31	.58
5	.33	.44
6	.28	.48
7	.39	.54
8	.40	.54
9	.31	.46

depending on the alignment of the question to best practice literature, were more likely to have overall higher trust alignment scores. The only scenario which did not have an index of 0.30 or higher was scenario six, but its value was 0.28. The Cronbach's alpha would decrease if seven of the nine scenarios were removed, and the Cronbach's alpha would remain the same if scenario two was removed and increase by 0.01 if scenario four was removed.

Table 4.7 shows the students' perceptions of trust for each scenario. The complete scenario can be found in Appendix B. For eight of the nine scenarios (all but scenario 8), over half of the students reported their trust in their teacher would change in accordance to their sense of fairness (scenarios which were considered fair would increase trust while scenarios which were considered unfair would decrease trust.) For scenario 8, 55% of the students felt using one type of assessment would decrease their trust in their teacher, while 50% of these students felt such practice was fair.

Table 4.7

Student Perceptions of Change in Trust Results Shown as a Percentage

Scenario Number	I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.
1	26.96%	26.40%	23.60%	9.64%	4.89%	8.52%
2	13.05%	21.57%	35.58%	13.74%	4.67%	11.40%
3	5.04%	8.45%	15.26%	24.11%	19.07%	28.07%
4	12.30%	16.94%	33.20%	15.98%	6.69%	14.89%
5	9.70%	14.96%	31.99%	17.59%	7.06%	18.70%
6	4.29%	7.47%	12.72%	20.33%	20.89%	34.30%
7	10.72%	11.84%	24.51%	18.11%	12.40%	22.42%
8	7.81%	13.21%	33.95%	19.32%	9.23%	16.48%
9	7.91%	11.44%	23.31%	24.86%	11.44%	21.05%

These percentages represent the students' change in trust in response to the different scenarios presented. However, in order to complete inferential data analysis, the results first had to be coded as continuous variables, as designated on the survey (e.g. "6" representing "I would completely trust my teacher" and "1" representing "I would completely not trust my teacher"). Then, the scores of any questions which deviated from best practice needed to be recoded. This would ensure better clarity in analyses, as a "6" would now represent a student's change in trust for an unfair scenario (as described by best practice literature) as "I would completely not trust my teacher." Similarly, a "1" in this same scenario would represent a student's change in trust for an unfair scenario as "I would completely trust my teacher.") For scenarios which were written as promoting best practice (as described in the research literature), these trust values did not

need to be reverse coded. Thus, a "6" would represent a student whose trust level on a fair scenario would change to "I completely trust my teacher" and a "1" would indicate "I completely do not trust my teacher" on an unfair scenario. Finally, each student's total score for fairness was created by adding all the student's reverse coded fairness scores; a trust total score was also created in a similar fashion.

The maximum score a student could receive for each variable was a fifty-four (nine scenarios with a maximum score of six). The maximum observed score was fifty-three for fairness, and a fifty-two for trust. The minimum observed score for both fairness and trust was a nine. Table 4.8 shows the means and standard deviations for both variables.

Table 4.8

Mean and Standard Deviations for Fairness Total Scores and Trust Total Scores

	Mean	Standard Deviation
Fairness	34.66	6.08
Trust	33.93	5.37

Hypothesis Testing: Trust Perceptions

Once the responses were reverse coded, the researcher conducted the various assumption tests for multiple regression. First, the test for independence of residuals was conducted using the Durbin-Watson statistic. With a value of 1.95 the null hypothesis that there was not independence was rejected and the assumption was allowed. The next assumption specifies a need for there to be

a linear relationship between (a) the dependent variable and each of the independent variables, and (b) the dependent variable and the independent variables collectively. Using the scatterplot depicted in Figure 4.1, the researcher detected a linear relationship existed, and the assumption was met.

Figure 4.1. Scatterplot of fairness and trust scores differentiated by gender.
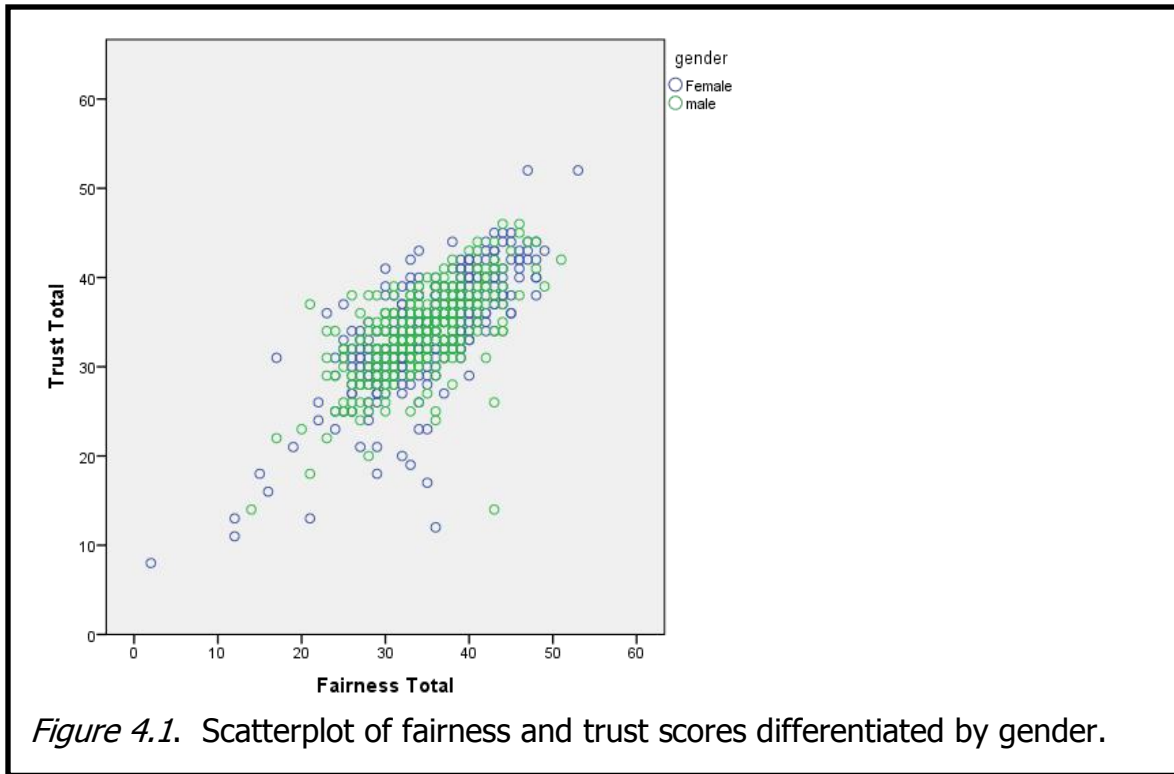
Next, the researcher needed to show the data met the assumption of homoscedasticity, which is where the variances along the line of best fit remain similar as you move along that line. Using a scatterplot of residuals and predicted values, the regression standardized residual values are approximately equally distributed for all values of the regression standardized predicted values, which shows the assumption of homoscedasticity is preserved. This scatterplot is shown in Figure 4.2.

Figure 4.2. Scatterplot for test of homoscedasticity
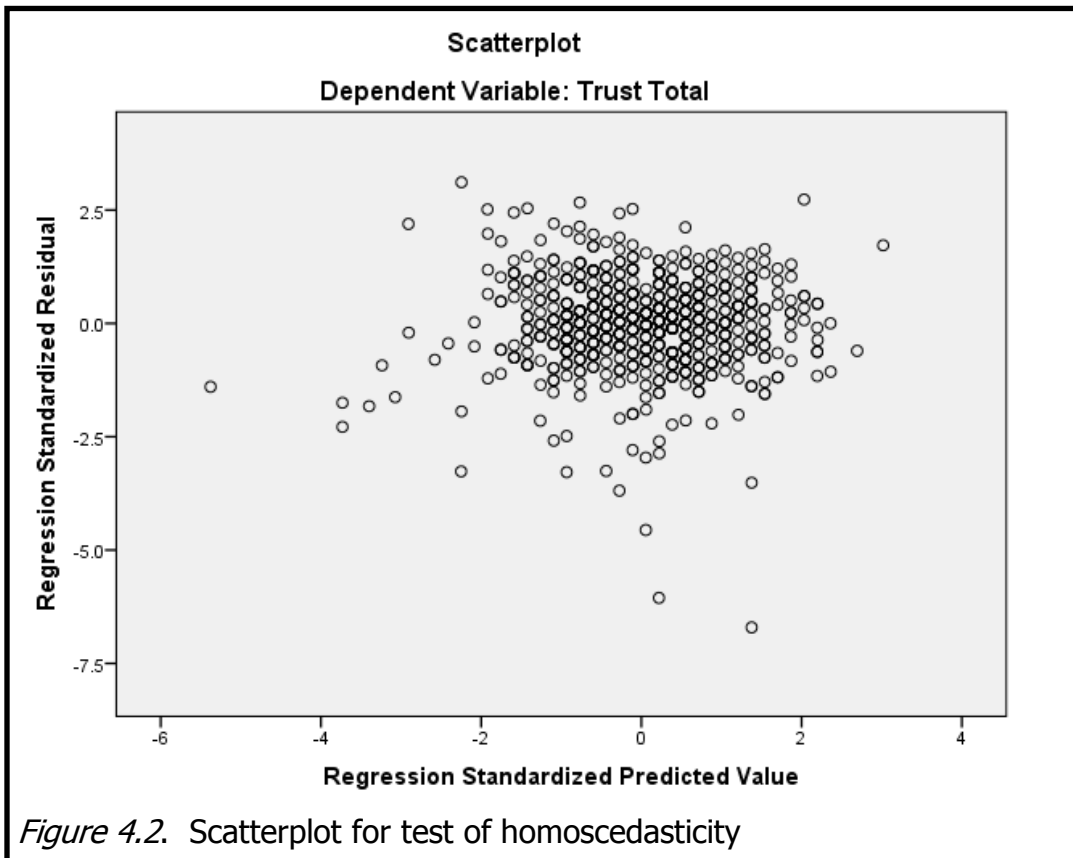
The next assumption that must be checked for is multicollinearity, which occurs when you have two or more independent variables that are highly correlated with each other. This can be tested using Variance Inflation Factors. Generally, if the value is less than 10, there is no evidence of multicollinearity. This data set had a value of 1.001, which allows the assumption of no multicollinearity to remain intact.
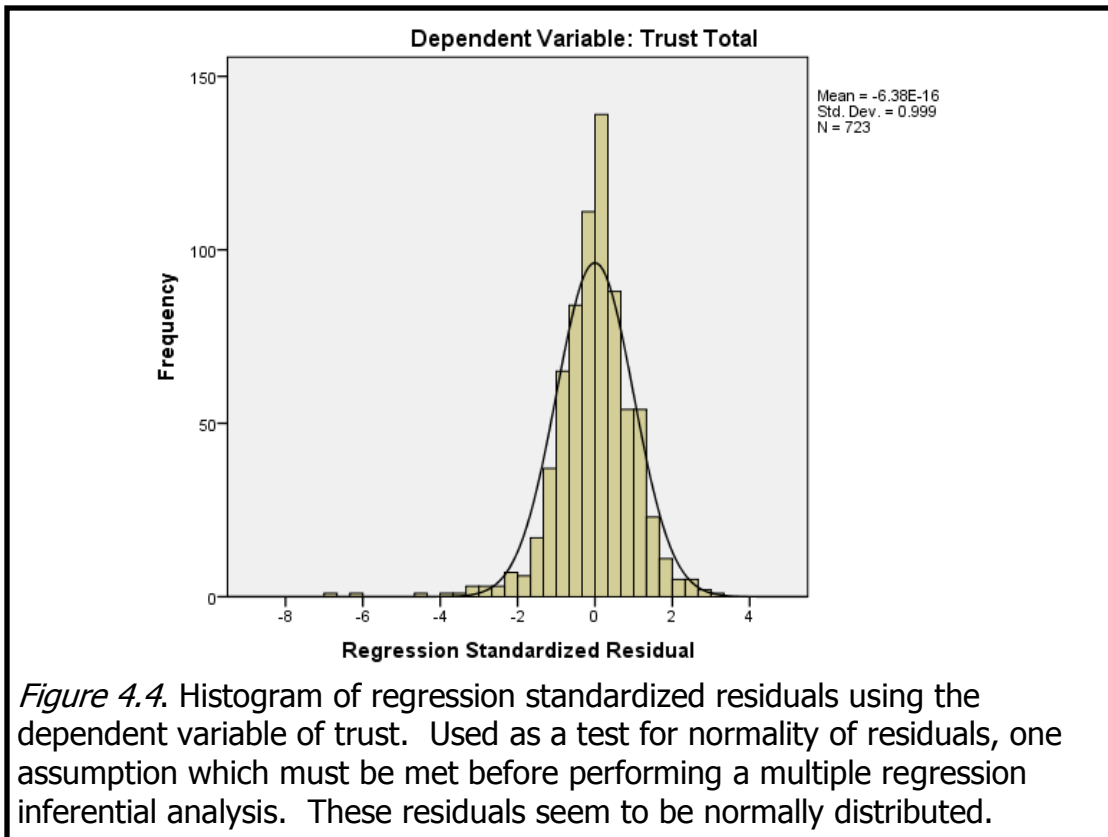
The next test for assumptions addresses outliers. In this data set, there were nine outliers, as identified by the Casewise Diagnostics in SPSS, which is shown in Figure 4.3. These cases were not deleted; however the analysis was run with and without the apparent outliers. The same conclusion was reached, regardless of the inclusion or deletion of the outliers.

58

Case Number	Std. Residual	Trust Total	Predicted Value	Residual
86	-3.26	20	32.24	-12.24
124	-3.69	19	32.87	-13.87
152	-4.56	17	34.14	-17.14
186	-6.06	12	34.77	-22.77
361	-3.27	13	25.28	-12.28
379	-3.28	18	30.34	-12.34
393	3.11	37	25.29	11.71
664	-6.71	14	39.21	-25.21
672	-3.51	26	39.21	-13.21

Figure 4.3. Casewise diagnostics results from multiple regression analysis. These nine cases were identified as outliers, and were removed from the data set. The dependent variable was the Trust Total Scores.

Finally, the last assumption involves ensuring the residuals are approximately normally distributed. Using a histogram (Figure 4.4), the researcher concluded this assumption was met.

Once determining the assumptions for multiple regression were met, the actual regression analysis was completed. The student's total Fairness Alignment Coefficient, which represented the student's alignment to best practices, and the student's gender were used as independent variables. The student's overall trust score, which represented the degree to which a child's trust changed positively for fair situations and negatively for unfair situations, was used as the dependent variable.

Figure 4.4. Histogram of regression standardized residuals using the dependent variable of trust. Used as a test for normality of residuals, one assumption which must be met before performing a multiple regression inferential analysis. These residuals seem to be normally distributed.

Using both fairness total and gender as predictors, the model yields an adjusted R squared value of .510, which indicates the model accounts for a little more than half of the variation in trust scores. Cohen's f^2 effect size for this R squared value is 1.0, which Cohen (1992) describes as a large effect size. The model is statistically significant in predicting trust from FAC and gender (p<.0005). In looking at the individual predictors, the FAC is a much stronger predictor, with gender not being a significant predictor (p=.954). Table 4.9 shows the coefficients and significance values for the model.

The following chapter includes some discussion of the outcomes from the various tests, interpretation of the findings as a whole, and implications for further analysis.

Table 4.9

Multiple Regression Coefficient Table with Trust Total as Dependent Variable

Model	Unstandardized Coefficients		Standardized Coefficients	Significance
	B	Std. Error	β	
Constant	11.98	.92		.00
FAC	.633	.02	.72	.00
Gender	.016	.28	.00	.95

Chapter Five

Discussion and Conclusion

Summary of the Study

The purpose of this study was to investigate student perceptions of

fairness in common classroom grading practices as well as change in trust of the

teacher after these practices. With classroom formative and summative

assessment, as well as grading policies, being a frequently discussed and

debated topic, and the lack of empirical evidence concerning these practices and

policies, this study informs the literature about one stakeholder's perceptions of

these topics. This study demonstrates that, generally, students do perceive

practices which align to the pre-existing literature as fair and practices which

deviate from the pre-existing literature as unfair, as identified in the mean values

for the fairness responses. Furthermore, the within-subjects ANOVA supported

the supposition that the students are more discriminatory in their perceptions

based on the type of fairness represented in the teacher's actions, as found in

the post-hoc examination of the FAC's for outcome, interactional and procedural

fairness.

The regression analysis in this study also demonstrated there was a

relationship between a student's fairness alignment to best practices (FAC) and

the student's change in trust in their teacher. Gender, which was also explored

in this same analysis, was not a significant predictor of trust levels. This result suggests that as students feel a teacher is acting more and more fairly, they will trust that teacher more. The inverse is also true: this study's results suggest that as students feel a teacher is acting more and more unfairly, they will distrust that teacher more and more.

Discussion of Results

This section provides a discussion of the results including the factor analysis, analysis of variance, and regression analyses. It also examines the similarity of results to those reported in the literature, potential factors contributing to the results, and challenges for future studies.

Fairness alignment and change in trust.

Descriptive statistics provided some evidence to support the hypothesis that students were able to identify situations that were written to deviate from best practice as more unfair and situations that were written to promote best practice as fair. Using within-subjects' analysis of variance, contrary to the null hypothesis, there were differences in the students' fairness alignment depending on the type of fairness presented in the scenario. Scenarios involving interaction fairness were different from both outcome and procedural fairness scenarios.

In the regression analysis, two factors were used as predictors in the students' change in trust: the student's alignment to fairness coefficient and the student's gender. Student's gender was found to not be a significant predictor

in change on trust. However, the more a student could identify fairness

scenarios correctly based on literature, the more the student's trust changed in

alignment to if the practice deviated or aligned to best practice.

Connections to the literature.

While there has been discussion of fairness in standardized testing,

fairness in classroom education has been limited to mainly university students.

The results of this analysis appear consistent with a result from one of those

studies. Gordon and Fay (2010) examined several hundred college students'

perceptions of fairness in their evaluative systems. They found that college

students believed interactional fairness (the way in which the professor treats

the students in terms of equality) to be more important than outcome fairness

(how teachers used assessment to create the course grade). In the current

study, the results from the Tukey post hoc analysis confirmed middle school

students also were more perceptive to interactional fairness promotion and

deviation than outcome and procedural fairness.

This study also contributed to the literature concerning trust and fairness

in gender. Previous studies (Baker, 2006; Howes et al., 2000; Hughes, Cavell, &

Wilson, 2001) found girls in pre-K to 6th grade to trust their teachers more,

regardless of the actions of the teacher. Boys in those grades trusted their

teachers less. However, this research on assessment practices found the gender

of the student was not a significant factor in predicting if the student's trust

would change based on the teacher's actions and the student's perception of those actions.

Implications of Results and Limitations

The study's primary purpose to discover if students' perceptions of assessment practices aligned to best literature was statistically inconclusive. While seven of the nine scenarios had a rounded mean of 4 (which indicated some alignment) or higher, the standard deviations for all of the scenarios were higher than 1.5, which allows for statistic uncertainty of this alignment. Perhaps the fairness question should not have been a Likert scale response, but a binary response ("Is it fair?" "Yes" or "No"). This would allow for a more precise interpretation of the results. However, it also may lead to less valid results in that fairness is not always seen as a binary issue.

Another possible limitation was the sample response rate. Although there were 738 responses, this only represented 34.5% of the total population of seventh grade students in the district. While the larger number of responses allowed for adequate power, there might still be some response bias. Because this survey was sent by email, these responses may represent students who are more technologically inclined, or who have more frequent access to their email. It was also sent from the seventh grade administrator for each school. This was done in an effort to increase participation (receiving an email from a known person rather than an unknown district office email address). However, students

who may have had a negative interaction with the assistant principal may have been less inclined to participate in the survey.

Recommendations

There are several recommendations for improving classroom assessment related to this study. First, student perceptions of fairness do appear to, generally, if not statistically, align with assessment best practices. Therefore, seventh grade teachers who desire to be considered fair by their students should reflect on their grading practices and move to align them to the best practices literature described in this study. By doing so, these teachers will likely not only be regarded as more fair, but also garner more trust. Trust in teachers increased as students were able to perceive fair and unfair situations, as defined by the assessment literature.

Upon this reflection, if a teacher finds many practices are out of alignment, that teacher may start with aligning any assessment practices that may impact interactional fairness first. This would include any assessment practices that influence the daily relationship between instructor and student, most likely formative assessment situations. This recommendation comes from the within-subjects' analysis of variance results, which found students' perception were aligned more closely to best practices when the scenario described interactional fairness.

The exploratory factor analysis also lends itself to recommendations. Students reacted in a similar manner to scenarios which appeared to have a latent variable which represented justice, empathy and control. Thus, this may suggest that teachers should attempt to view their assessment practices from these standpoints. Reflective questions may include "Would my students see this practice as a means to employ justice in the classroom?" or "Does performing this practice make me seem more empathetic?" or "Am I using this practice as a means to control my students?" Answering in the affirmative for any of these questions does not necessarily mean a teacher should change the assessment practice in question. The research would suggest the teacher should align the practice to the literature. However, if the teacher answers one of these reflective questions in the affirmative, that teacher may also wish to address the reasoning behind the decision when reviewing assessment policies with students. For example, a teacher may believe giving a zero to students who do not turn in assignments is a means to employ justice to those who fail to do their work or who are forgetful. Students also see it in this respect. However, giving zeros deviates from best assessment practice literature. So, if a teacher decides to not give zeros, but instead have students stay and make up work, that teacher may want to describe the rationalization behind this decision so students no longer see the zero as a means to justice.

Implications for Future Research

First, this study was specifically performed with middle school students due to their unique development at this age. However, expanding it to elementary and high school students would provide a larger picture of the relationship between best practice scenarios, fairness, and trust. Yet, there are roadblocks to this idea, as some of the scenarios were written with middle school classes in mind.

To investigate if the latent-factors identified by the research are valid, the utilization of qualitative methods might be helpful. Interviewing students about the scenarios, and why they chose the level of fairness may give insight into if the identified factors of justice, empathy, and control are supported. However, the type of research may be cost prohibitive for the number of students the researcher would need to interview, as well as the time from class each child would miss.

Also, only nine scenarios were developed for this survey to restrict time needed to take the survey to a homeroom period. If there had been more questions which were written for each best practice, alignment, and type of fairness, we would expect the results to be more reliable. However, more scenarios would mean more time needed to take the survey.

Additionally, while this study unsuccessfully attempted to employ gender as a statistically significant predictive variable in the relationship between fairness and trust, there may be another variable which may yield significant

results. During the focus group, occasionally the students would want to comment on if they felt a particular scenario was fair. The researcher reminded them that this was not their task, and guided them toward the questions on the protocol sheet. However, the researcher noticed a pattern during these off-topic discussions. Students who were typically compliant students found the scenarios in which a student received a worse grade due to non-compliance as fair. For example, two students agreed with a third student's statement concerning the reading log not being turned in, "That student had plenty of chances, she deserves that zero" (Anonymous, 2015). Similarly, when discussing the student who threw away his assignment accidentally, these same students felt "giving him a second chance is wrong. He needs to be more organized" (Anonymous, 2015). After speaking to the students' teachers briefly, these three students are all extremely organized, work promptly in class, and rarely choose to not follow school rules. Therefore, it may be interesting to examine if there is a link between students who follow the typically compliant student model have different perceptions of fairness in terms of assessment.

Finally, standardized assessments are a major part of the educational accountability landscape. They influence school and district ratings and teacher contract status. Thus, studying classroom assessment practices as they relate to standardized test scores may be of interest. Since there is cursory evidence that students can perceive fair and unfair classroom assessment practices, it may be interesting for students to give their current teachers a fairness rating, based on

the literature. Then, a researcher could examine if there is a relationship between standardized test achievement and teachers' fairness rating.

Conclusion

As a result of this study, there is some evidence which supports students can identify practices which promote best assessment practice and practices which deviate from best assessment practices. Students are more inclined to be in alignment to best assessment practice when interactional fairness is at stake. Students whose fairness perceptions were more aligned with best practice literature were more likely to gain or lose trust in their teacher, depending on if a scenario promoted or deviated from best practice. Thus, educators of middle school students might want to strive to align their assessment practices to best practice if for no other reason than their students will generally be able to perceive when they do or do not. Not only will the students' perceptions of fairness change as teachers align their practices to literature, but so will their levels of trust.

References

Alkharusi, H. (2008). Effects of classroom assessment practices on student's achievement goals. *Educational Assessment, 13*, 243-266.

Alkharusi, H. (2010). Teachers' assessment practices and students' perceptions of the classroom assessment environment. *World Journal on Educational Technology, 2*(1), 27-41.

American Educational Research Association. (2014). *Standards for educational and psychological testing.* Washington, D.C.: American Educational Research Association.

Anonymous. (2015). [Comments and notes from educator survey and focus group.] Unpublished raw data.

Assessment methods: A set of four options. (2011). Pearson Assessment Training Institute. Retrieved from http://www.sllboces.org/site/handlers/filedownload.ashx?moduleinstancei d=4363&dataid=3417&FileName=AssessmentMethodsASetofFourOptions. pdf

Baker, J. A. (2006). Contributions of teacher-child relationships to positive school adjustment during elementary school. *Journal of School Psychology, 44*, 211-229.

Black, P., & Wiliam, D. (1998). Inside the black box: raising standards through

    classroom assessment. *Phi Delta Kappan, 80*, 139–144.

Brookfield, S.D. (2000). *The skillful teacher*. San Francisco: Jossey-Bass

Brookhart, S. M. (1993). Teachers' grading practices: meaning and values.

    *Journal of Educational Measurement, 30(*2), 123-142.

Brookhart, S. M., & DeVoge, J. G. (1999). Testing a theory about the role of

    classroom assessment in student motivation and achievement. *Applied*

    Measurement in Education, 12, 409–425.

Brown, J. B. (1999). The use of focus groups in clinical research. In (Eds.)

    Crabtree, B. F., & Miller, William L. *Doing qualitative research* (2nd ed.)

    (pp. 109-124). Thousand Oaks: Sage.

Bryk, A.S., & Schneider, B. (2003). Trust in schools: A core resource for school

    reform. *Educational Leadership, 60(*6), 40-45.

Cameron, J., & Pierce, W. (1994). Reinforcement, reward, and intrinsic

    motivation. *Review of Educational Research, 64(*3), 363–424.

Carlson, L.A. (2003.) Beyond assessment to best grading practice: Practical

    guidelines. In Wall, J.E. & Walz, J.R. (Eds.) *Measuring up: Assessment*

    issues for teachers, counselors, and administrators (507-515). Austin, TX:

    PRO-Ed Inc.

Christine, D. and Viner, R. (2005). ABC of adolescence: Adolescent

development. *British Medical Journal, 330*, 301-304. Retrieved from

http://peds.stanford.edu/Rotations/adolescent_medicine/documents/ABCs

ofAdolescentDevelopment.pdf

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 122, 155–159.

Doherty, G.D. (2008). On quality in education. *Quality Assurance in Education,*

16(3), 255-265.

Eccles, J.S., & Midgley, C. (1989). Stage/environment fit: Developmentally

appropriate classrooms for early adolescents. In R.E. Ames & C. Ames

(Eds.), *Research on motivation in education* (pp.139-186). San Diego,

CA: Academic Press.

Fink, A. (2003). *How to ask survey questions*. Thousand Oaks, CA: Sage

Publications.

Fink, A. (2006). *How to conduct surveys: A step-by-step guide.* Thousand Oaks,

CA: Sage Publications, Inc.

Fowler, F.J., Jr. (2002). *Survey research methods.* Thousand Oaks, CA: Sage

Publications, Inc.

Gordon, M.E., & Fay, C.H. (2010). The effects of grading and teaching practices

on students' perceptions of grading fairness. *College Teaching, 58*, 93-98.

Green, S.K., Johnson, R.L., Kim, D., & Pope, N.S. (2007.) Ethics in classroom

assessment practices: Issues and attitudes. *Teaching & Teacher*

Education 23(7), 999-1011.

Green, S.K., & Johnson, R.L. (2010.) *Assessment is essential.* New York, NY:

  McGraw-Hill.

Gronlund, N.E. (2006). *Assessment of student achievement.* Boston, MA: Allyn

  and Bacon.

Gulikers, J.T.M., Bastiaens, T.J., Kirscner, P.A., & Kester, L. (2008). Authenticity

  is in the eye of the beholder: Student and teacher perceptions of

  assessment authenticity. *Journal of Vocational Education & Training,*

  60(4), 401-412. doi: 10.1080/13636820802591830

Guskey, T.R. (2009). *Practical solutions for serious problems in standards-based*

  grading. Thousand Oaks, CA: Corwin Press.

Hill, IV., G.W., & Zinsmeister, D. D. (2012.) Becoming an ethical teacher. In

  Buskist, W. & Benassi, V.A. (Eds.) *Effective college and university*

  teaching: Strategies and tactics for the new professoriate (125-133). Los

  Angeles: SAGE Publications, Inc.

Holmes, L.E., & Smith, L.J. (2003). Student evaluations of faculty grading

  methods. *Journal of Education for Business 78*(6), 318-323.

Howes, C., Phillipsen, L. C., & Peisner-Feinberg, E. (2000). The consistency of

  perceived teacher-child relationships between preschool and

  kindergarten. *Journal of School Psychology, 58*(2), 115-132.

Hughes, J. N., Cavell, T. A., & Wilson, V. (2001). Further support for the

  developmental significance of the quality of the teacher-student

  relationship. *Journal of School Psychology, 39*(4), 289-301.

James, A. R., Griffin, L., & Dodd, P. (2009). Perceptions of middle school

    assessment: An ecological view. Physical Education and Sport

    Pedagogy, 14, 323–334. doi:10.1080/17408980802225792

Johnson, R.L., & Morgan, G. (in press). *Survey scales: The essentials*. New

    York: Guilford Publications.

Kuhs, T.M., Johnson, R.L., Agruso, S.A., & Monrad, D.M. (2001). *Put to the test:*

    Tools and techniques for classroom assessment. Portsmouth, NH:

    Heinemann.

Martone, A., & Sireci, S.G. (2009). Evaluating alignment between curriculum,

    assessment and instruction. *Review of educational research, 79*(4), 1332-

    1361.

Marzano, R.J. (2000). *Transforming classroom grading*. Alexandria, VA:

    Association for Supervision and Curricular Development.

McMillian, J.H., Myran, S., & Workman, D. (2002, March/April). Elementary

    teachers' classroom assessment and grading practices. *The Journal of*

    Educational Research, 95(4), 203-213.

McMillan, J. H. (2004.) Classroom assessment: Principles and practice for

    effective instruction. Boston: Pearson.

McNeil, L. *Contradictions of school reform: Educational costs of standardized*

    testing. New York: Routledge.

Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test

    scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues*

    and Practice, 8(1), 14–22.

Nicholls, M.E.R., Orr, C.A., Okubo, M., and Loftus, A. (2006). Satisfaction

    guaranteed: The effect of spatial biases on responses to Likert scales.

    Psychological Science, 17, 1027-1028.

Pope, N., Green, S. K., Johnson, R.L., & Mitchell, M. (2009). Examining teacher

    ethical dilemmas in classroom assessment. *Teaching and Teacher*

    Education, (25), 778-782.

Popham, W. J. (1991). Appropriateness of teachers' test preparation practices.

    E*ducational Measurement: Issues and Practice*, 10(4), 12–15.

Preston, C.C., & Colman, A.M. Optimal number of response categories in rating

    scales: Reliability, validity discriminating power, and respondent

    preferences. *Acta Psychologica, 104,* 1-15. Retrieved from

    https://www.academia.edu/160132/Preston_C._C._and_Colman_A._M._20

    00_._Optimal_number_of_response_categories_in_rating_scales_Reliabilit

    y_validity_discriminating_power_and_respondent_preferences._Acta_Psyc

    hologica_104_1-15

Raebeck, B. (1993). *Exploding myths, exploring truths: Humane, productive*

    grading and appropriate grouping in the quality middle school. Paper

    presented at the annual conference and exhibit of the National Middle

    School Association, Portland, OR.

Rodabaugh, R.C. (1996.) Institutional commitment to fairness in college

teaching. *New Directions for Teaching and Learning, 1996*(66), 37-45.

Rodabaugh, R. C., & Kravitz, D. A. (1994.) Effects of procedural fairness on

student judgments of professors. *Journal on Excellence in College*

Teaching, 5(2), 67-84.

Ross, A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Influences on student

cognitions about evaluation, assessment in education: Principles, policy &

practice, 9:1, 81-95. doi: 10.1080/09695940220119201

Sadler, D. R. (1989). Formative assessment and the design of instructional

systems. *Instructional Science, 18* (2), 119–144.

Sloan, W.M. (July 2012). What is the purpose of education? Education Update,

54(7). Retrieved from

http://www.ascd.org/publications/newsletters/education-

update/jul12/vol54/num07/What-Is-the-Purpose-of-Education¢.aspx

South Carolina Department of Education. (2014). Poverty index [Data file].

Retrieved from http://ed.sc.gov/data/report-

cards/2014/documents/2014_Report_Card_Poverty_Index.xls

Stiggins, R. (2008). *An introduction to student-involved assessment for learning*.

Upper Saddle River, NJ: Pearson Education Inc.

Stranahan, H.A., Borg, J.R., & Borg, M.O. (2009). School grades based on

      standardized test scores: Are they fair? *Journal of Academic and Business*

      Ethics, 1(1). Retrieved from

      http://www.aabri.com/manuscripts/08006.pdf

Strauss, A. & Corbin, J. (1990). Basics of qualitative research: Grounded theory

      procedures and techniques. Newbury Park, CA: Sage Publications.

Thapa, A., Cohen, J., Guffey, S., & Higgein-D'Alessandro, A. (2013). A review of

      school climate research. Review of Educational Research, doi:

      10.3102/0034654313483907.

Tierney, R.D. (2014). Fairness as a multifaceted quality in classroom

      assessment. *Studies in Educational Evaluation*, *43*, 55-69.

Tierney, R.D., Simon, M. & Charland, J. (2011). Being fair: Teachers'

      interpretations of principles for standards-based grading. *The Educational*

      Forum, 75, 210-227. doi: 10.1080/00131725.2011.577669

Trust. (2015). In *Oxford Dictionaries*. Retrieved from

      http://www.oxforddictionaries.com/us/definition/american_english/trust

Tyler, T.R., Boeckmann, R.J., Smith, H.J., & Huo, Y.J. (1997). Social justice in a

      diverse society. Boulder, CO: Westview Press.

Velasquez, M., Andre, C., Shanks, T., Meyer, M.J. (1996). Thinking ethically: A

   Framework for moral decision making. *Issues in Ethics, 7*(1). Retrieved

   from http://www.scu.edu/ethics/publications/iie/v7n1/thinking.html

Vieno, A., Perkins, D.D., Smith, T.M., & Santinello, M. (2005). Community on

   school: A multilevel analysis. *American Journal of Community Psychology,

   36*(3-4), 327-341. doi: 10.1007/s10464-005-8629-8

Wang, M., & Holcombe, R. (2010). Adolescents' perceptions of school

   environment, engagement, and academic achievement in middle school.

   American Educational Research Journal, 7(3), 633-662. doi:

   10.3102/0002831209361209

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived

   pedagogical caring. *Journal of Educational Psychology, 89*(3), 411-419.

Wentzel, K.R. (1998). Social relationships and motivation in middle school: The

   role of parents, teachers and peers. *Journal of Educational Psychology,

   90*(2), 202-209.

Whitley, Jr., B.E., Perkins, D.V., Balogh, D.W., Keith-Spiegel,P., & Wittig, A.F.

   (2000). Fairness in the classroom. *Association for Psychological Science's

   Observer, 13*(6). Retrieved from

   http://www.psychologicalscience.org/teaching/tips/tips_0700.html

Wiggins, G., & McTighe, J. (1998). Understanding by design. Alexandria, VA:

   Association for Supervision and Curriculum Development.

Wormeli, R. (2006). Fair isn't always equal: Assessing and grading in the

    differentiated classroom. Portland, ME: Stenhouse Publishers.

Wren, J., Sparrow, H., Northcote, M., & Sharp, S. (2009). Higher education

    students' perceptions of effective assessment. *The International Journal*

    Of Learning, 15, 11-23.

Appendix A-Focus Group Script and Questions

I am studying seventh graders thoughts on how teachers assess students and grade their work. I am specifically looking at if the actions of the teacher are fair.

1. When I say fair in relation to assessment and grading, what things do you think of?

I am going to describe some scenarios to you. Think of them like stories. They are written for you on the sheets I am giving you now. After each, I am going to ask you three questions:

a. *Was it clear?* What is happening in the scenario?

b. *Is it possible?* I want to know if you could imagine this actually happening in middle school. It might not have happened to you, but you could see how it might happen to someone in middle school.

c. *How can I make it better?* This is exactly what it sounds like. If you think it is clear and possible, there might be no need to make it better, but we'll chat to make sure.

(Researcher then passes out scenarios and discusses each one with the three questions listed above.)

Appendix B-Survey and Scenarios

Hello,

My name is Cori Jimenez, and I am a student at USC. I also work for the district with their data team. I am studying what students, like you, think about how teachers give feedback to students and how they assign grades. I am also interested in what might help build trust between a teacher and student. I hope my research will make our district a better place for us all to learn.

On the next few pages, you will be asked to read some scenarios, or stories. The actions might have happened to you, or a friend of yours, or might have never happened to you or anyone you know. That's OK. I want you to pretend you are in the classroom where this happened. Then answer two questions: one about fairness and one about trust. There are no right or wrong answers, no trick questions. Your opinions are valid, no matter what they are. Just take some time and answer honestly, that's all I ask.

You do <u>not</u> have to complete this survey. It is not tied to any grade, and you will receive no physical reward or any punishment for completing it. You will know that you are helping all the future middle schoolers in the district. I do ask for your school email, but I will keep your responses confidential. This means I

82

will not share what you decide with other people. I will be the only one who

knows. Thanks a ton!

Scenario 1: (Deviates from best assessment practice)

Your science teacher required each student to complete a science experiment. You needed to go through the science process (pick a question, research your topic, make a hypothesis, etc.) Then, you needed to create a poster to show what you learned at the end of the experiment. You had to write everything using your handwriting. Your teacher gave you a rubric that included exactly how she was going to grade the experiment, and included a section on neatness.

One of your classmates is really great at science. She knows every answer, and is always performing mini-experiments. She had the most impressive experiment you have ever seen, and even your teacher said she had never seen a better experiment. When everyone got their grades back, you noticed your friend had a B. When you asked her about it, your friend said the teacher had taken away points only because her handwriting is awful, no matter how hard she tries.

How fair is it for your classmate to get her science grade lowered because of her handwriting?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

84

Scenario 2: (Deviates from best assessment practice)

One of your teachers is trying to motivate everyone to do their best. One way he is trying to increase the students' motivation is by handing back tests to students in order of how well they did. Students with the highest grades get their tests back first, and students with the lowest grades get their tests back last.

How fair is it for the teacher to hand back tests in order from highest to lowest?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 3: (Promotes best assessment practice)

Your class has been asked by your Spanish teacher to create a report on a country where most people speak Spanish. You have had two weeks to work on it in class. A classmate of yours accidentally throws out his report when he was cleaning out his book bag. The next day, your Spanish teacher asks for the reports to be turned in. Your classmate realizes what happened, and tells the teacher. Your classmate thinks he will get a zero, but, instead, she tells him he has to work during lunch and/or after school until he does the report again.

How fair is it for the teacher to allow him redo the assignment?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 4: (Deviates from best assessment practice)

At the beginning of each math class, you review that night's homework. More and more of our classmates are not doing their homework. Your math teacher is frustrated. She reminds everyone that homework is practice, and practice makes perfect.

Meanwhile, you have just started a new unit in math. It is one a topic that is pretty new to you and really challenging. The homework was hard, but you tried your best.

The next day, your teacher says she is collecting homework and will be grading it. The grade will be a percentage of how many questions you got right.

How fair is it for the teacher to grade this homework assignment?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 5: (Promotes best assessment practice)

You have been studying writing for information. Your reading teacher puts everyone in groups. Each group has to write an article and then work together to create a newspaper. You notice a female in another group is working, but whenever someone in her group asks her a question, she doesn't look at the person and doesn't stop her work. The group complains to the teacher, who asks the female to be nicer.

The groups publish their newspapers, and then review the other groups' newspapers. You notice her article is really well written. But you wonder if her bad attitude made her lose points. So you ask her. She says she got full points.

How fair is it for the female to get full points?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 6: (Promotes best assessment practice)

You have been learning about the Industrial Revolution in social studies. Your teacher thinks it is very important for you to know how to read a timeline, and figure out how much time passed between dates. Your class has been working on that skill over and over again, subtracting the younger date from the older date.

Your teacher notices some of the class is struggling with subtraction. Every time they set it up correctly, but make an error on the calculation. When the teacher notices your classmates struggling with the calculations, she gives them a calculator.

The day of the test, your teacher sets out a basket of calculators. She tells the class they are welcome to use a calculator on any part of the test.

How fair is it for the students to get to use a calculator?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 7: (Deviates from best assessment practice)

In reading class, your teacher has you read 100 minutes during the week and write a summary about what you read. It is due every Monday.

A classmate of yours is really disorganized, but a great reader. She always has a book out at lunch. But, she usually forgets to hand her reading log in Monday. Your teacher has been giving her warnings when she turns it in late.

Finally, one week, your teacher finally does what she has been warning: she refuses to accept the late reading log. The female gets a zero for that week.

How fair is it for the teacher to give the female a zero for late work?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 8: (Deviates from best assessment practice)

Your social studies teacher loves projects. And she makes really good ones, too. You have had to do things like become an historical figure for a museum, make laws like an emperor, and debate if the Earth revolves around the sun. You love it!

However, you have a classmate who is shy. He doesn't like to talk in front of people, and he had a hard time during all of those projects. Plus, he's not real creative, so coming up with ideas for some were hard. But, the teacher has told everyone that in real life there are no multiple choice tests, so everyone needs to learn how to make projects now.

How fair is it for the teacher to only assign projects?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.

Scenario 9: (Promotes best assessment practice)

Learning a new language is hard! All these new vocabulary words, and making your mouth move a new way can be tiring. Your French teacher is always having you try to say things, even if you aren't sure. Then he gives you pointers on how to make it sound better next time. If you don't know a word, you can just ask and your teacher tells you.

Your teacher sits down with everyone once a week and speaks with them for two minutes. You get a grade on how well you used the vocabulary from the week and formed the new words. You never get graded on your daily work, even if it's really good.

How fair is it for the teacher to not grade your daily work and only grade your weekly work?

1	2	3	4	5	6
Completely unfair	Somewhat unfair	A little unfair	A little fair	Somewhat fair	Completely fair

If this happened in your class, how much would it change your trust in your teacher?

1	2	3	4	5	6
I would not trust my teacher anymore.	I would trust my teacher a lot less.	I would trust my teacher a little less.	I would trust my teacher a little more.	I would trust my teacher a lot more.	I would trust my teacher completely.