

Disparities in Judgment: A Case of Concern in Testing and Assessment

Derrick D. Davis, *Tuskegee University*

Abstract

Without question, faculty (regardless of discipline) should be equipped with the necessary skills to assess students fairly and ethically. This study focuses on the central and prevailing importance of faculty judgment and how that judgment (or lack thereof) influences perceptions related to ethics and assessment of students. The study outlines the results from a scenario-based survey (adopted by Green, Johnson, Kim & Pope, 2007). The survey was given to in-service teachers across a K-12 school district and measures whether or not teachers either strongly agreed or strongly disagreed when requested to judge the ethics of various assessment practices presented. Findings revealed that sharp divisions exist among educators related to assessing whether something is or is not an ethical assessment practice.

Keywords: evaluation, assessment, ethical, faculty, judgment

If one is looking for a “flashpoint” in education, it is the assigning of grades and the controversy that surrounds the ethics related to the assessment of students. In large part, assessment remains the one area where, despite a basic foundation within educational preparation programs promoting summative and formative assessment, few “guidelines” are designed that focus on the principles that should guide “ethical assessment practices” (Green, Johnson, Kim & Pope, 2007, p. 999; Popham, 2017). This lack of guidance within teacher preparation programs has been attributed to unethical grading practices that can either largely go unchecked or (once surfaced) has led to community backlash and legal woes for educational entities.

Further, controversial policies like “no-zero” policies (which routinely “ban [faculty from issuing] grades of less than 50%” are raising ethical concerns as well (Fink 2018; Walker, 2016, para.1). Such controversial ethical issues surrounding grades have led to firings of teachers for refusing to comply or resignations altogether (Fink 2018; Walker, 2016,). Indeed, in 2018, an eighth-grade Social Studies teacher in St. Lucie, Florida, reached internet fame and thrust her school district into controversy by posting her final message to students on Facebook (after being fired) that read: *“Bye Kids. Mrs. Tirado loves you and wishes you the best in life! I have been fired for refusing to give you a 50% for not handing anything in”* (Fink, 2018, para. 1). Equally important, other ethical issues have surfaced related to teachers lacking sound judgment in grading altogether. For example, a graduate student at Lehigh University in Pennsylvania (after receiving a “C” in a course that led to her dismissal from her program), sued when it surfaced that she he was given a zero for classroom discussion due to her behavior during classroom discussions (Zotter, 2019). Following years of litigation, the appellate panel stated that “While another instructor might have given a more lenient grade than zero for these infractions, we are reluctant to make a judicial determination to overturn [the] grade” (Zotter, 2019, para. 6). Despite the school’s legal victory, in this case, the time and expense of such litigations are costly and can damage the reputation of the institution. Guidelines and professional development for instructors would help to avoid such “follies”, improve ethical assessment practices, and prevent costly litigation.

Indeed, similar grade challenges, like the University of Texas San Antonio Student Government Association voting that its grading system was flawed, are growing (Davila, 2018; Rhor, 2012; Sloan, 2014; Zaretsky, 2014). Faced with growing grade challenges, the Texas Supreme Court, ruling on a separate case, stated that courts should only intervene in grade matters when the “school did not exercise professional judgment” (Feldman, 2019, para. 3). Hence, professional judgment is key in determining whether or not evaluation and assessment are fair and just, and educational entities are vulnerable if this is not addressed via professional training of its faculty.

Purpose and Significance

Professional judgment is key and significant in determining whether or not evaluation and assessment are fair and just. If unaddressed, education suffers, for the assessment of students has major implications and directly impacts education at its very core. Thus, its vital that faculty perceptions regarding what is or is not ethical as it relates to assessment are not only examined, but measured to determine if gaps exist. Hence, the central purpose of the study was to examine whether or not educators recognize and understand what is or is not ethical within assessment parameters and practices. Specifically, the study measures the degree in which agreement or disagreement among licensed K-12 educators is reached after examining various assessment scenarios to determine if ethics were violated. This study is significant for it is designed to

illuminate whether or not there are gaps in educators' perceptions related to ethical or unethical assessment practices, and if so, what should school districts and post-secondary entities do in order to combat such discrepancies and improve validity in assessment. Daily, K-12 teachers are faced with having to make professional judgments as it relates to assessment practices and oftentimes these professional judgments are questioned or disputed, and in some cases, result in litigation. Thereby, identify gaps in understanding of what is ethical or unethical practices in assessment, and illuminating the need for continuous professional development, is important and serves a key significant purpose in improving assessment outcomes.

Review of Literature

Efforts to guide ethical practices in assessment have been outlined in various publications to include: The Standards for Educational and Psychological Testing jointly published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (2014). Other publications include the Standards for Teacher Competence in the Educational Assessment of Students published by the American Federation (1990), and the Student Evaluation Standards outlined by the Joint Committee on Standards for Education (JCSEEE) which was published in 2003. However, despite these combined efforts, research shows that publications addressing the matter are not concrete or specific enough and do not render situational specific guidance for educators to properly vet situations and apply specific ethical standards to common scenarios (Fan, Johnson, Liu, Zhang, Liu & Zhang, 2019; Gipps, 1994; Green et al., 2007).

In addition, textbooks have also attempted to outline specific "ethical principles" that can serve to "guide ethical judgments related to assessment"; however, they do not provide essential guidance in response to how educators can make better "day-to-day judgment" altogether (Brookhart & McMillan, 2019; Fan et al., 2019; Feldman, 2018; Green et al., 2007, p. 1000; Heritage & Harrison, 2019; Popham, 2017). This deficit in the field of education, then, leads to ambiguity and lack of commonly established principles and core guidelines for educators across the spectrum to apply as judgment calls are being made. As such, K-12 school districts and post-secondary institutions rely on their own internal practices and policies to guide grading procedures. In doing so, there is no "common thread" within educational entities that serves as the "guiding North" that conjoins ethics and assessment in education. Educators are on their own, then, to figure out this connection, opposed to having standard ethical assessment practices that they can leveraged. As a result, school districts and post-secondary institutions are subject to increase lawsuits, firing and resignations related to unfair and unjust grading practices (Fink, 2018; McParland, 2020; Sloan, 2014).

Due to the lack of standardized ethical practices that should be adopted and implemented across K-12 and post-secondary education, "little research exists about the extent in which educators agree about the ethicality of student evaluation practices" altogether (Green, Johnson, Kim & Pope, 2008, p. 520). This is likely due to the fact that educators are not trained in this realm as undergraduates, and there is not generally agreed upon consensus about assessment practices and the ethics that surround them. This "lack of agreement [further] highlights the need for an overarching ethical framework from which to develop the capacity to make judgments about ethical assessment practices" (Gipps, 1994; Green et al., 2007, p.1000; Popham, 2000; Rasooli et al., 2018). And while some researchers promote the idea of establishing "ethical frameworks" that will help with "self-regulatory" guidelines for educators, to date, this has not been implemented

(Gipps, 1994; Green et al., 2007, p.1000). As such, there is no general consensus that guides educators—no litmus test—that helps to determine the ethicality of decisions reached related to assessment, yet there needs to be (Green et al., 2007).

Some researchers argue that codes should be established while others support the reliance on broader ethical principles like “Do No Harm” that Taylor and Nolen (2005) established that has been liberally applied as a guiding ethical principle, or the adage, “Treat others as you would have them to treat you”. However, such principles do not always work, for in applying judgment one has to have some formal training or expertise in the subject which many educators lack (Brookhart & McMillan, 2019; Fan et al., 2019; Feldman, 2018; Green et al., 2007, p. 1001). In examining the literature, it underscores that educators are lacking the “knowledge base” and may be ill “equipped” in making ethical judgments related to assessment practices largely due to the lack of formalized training in the area of assessment (Brookhart & McMillan, 2019; Fan et al., 2019; Feldman, 2018; Green et al., 2007). This lack of training and broader tentative understanding of the ethics surrounding assessment lend itself to educators making poor judgment calls which propels ethical dilemmas in the classroom related to assessment practices. Further, there is limited research on how educators actually apply their judgment as it relates to standing principles related to ethical assessment practices. Accordingly, there appears to be a gap between the standards, codes and practices related to ethical assessment practices and whether or not these concepts are applied when educators are faced with ethical dilemmas in the classroom. Research is limited in this area.

This lack of training and application of core standards has not only created a dilemma for the educators themselves, but for education entities as well. K-12 school districts and post-secondary institutions work to offer solid curriculum that supports student learning and are equally committed to providing fair and equitable assessment practices. Indeed, the importance of fair testing (in the classroom) cannot be understated, especially given the recent shift away from standardized testing. Actually, even before the “College Admissions Scandal” shocked the world in 2019 leading to several arrests and imprisonments, high stakes testing, which used to be considered the gold standard to measure and assess student learning, was being questioned (Tierney, 2014,). In fact, for much of the “20th Century”, high stakes testing was the mainstay, for it was widely accepted that standardized testing was objective, opposed to subjective (Tierney, 2014, p. 55). Yet, “[a]s the century turned, shifting social ideals, evolving ideas about the nature of knowledge, developments in understanding human learning and rapid technological advancements change the educational landscape” (Tierney, 2014, p. 55). This shift, then, moved conventional thought away from an over reliance on high stakes testing and shifted the emphasis to quality assessment and testing in the classroom (Johnson, Liu & Burgess, 2017; Pascal & Bertram, 2016; Rasooli, Zandi, & DeLuca, 2018; Stern, 2017; Widiastutu, 2018). Hence, the idea that if the shift has moved to a higher reliance on measuring student learning via assessment tools (e.g. formative, summative), and if high stakes tests like the ACT and SAT are now in question, it behooves the education community to address the reliability of its assessment tools and to better train and prepare educators on the ethics, “validity, reliability and fairness” of the assessments tools they produce (American Education Research Association [AERA], American Psychology Association [APA], and the National Council on Measurement in Education [NCME] (Tierney, 2014, p. 55).

Methodology and Approach

In order to best analyze teacher perceptions related to whether various assessment practices were ethical or not, a web-based survey (adopted and validated from a previous study that sought to investigate the same phenomena) was used (Green et al., 2007). The research design, Quantitative Analysis, was determined to be the best method and approach for this study, for the design enables the researcher to not only collect data numerically but also measure perceptions specific to various demographic variables (across the study). According to Creswell (1994), quantitative research enables the researcher to better investigate phenomena. It achieves this end by collecting quantitative data that are then analyzed using mathematically rooted statistical approaches. Specific to this study, the researcher is attempting to assess teacher perceptions about ethical or unethical behavior regarding assessment and evaluation practices (phenomena) and assess whether or not those perceptions align (agree) or do not align (disagree) over various demographic variables to include: gender, degree status, grade level taught, and years of experience. As noted, quantitative research, then, is essentially about collecting numerical data to explain a particular phenomenon. In this case, a survey was used to collect in-service teacher responses from various scenarios that centered around ethical or unethical assessment practices. The scenarios focused on specific assessment practices and asked teachers to read select scenarios and judge the outlined practices as either “ethical or unethical.” The study aimed to identify how much agreement or disagreement existed among k-12 educators as it related to the determination of whether or not a practice was ethical or not.

Framework of Research and Research Question

This study “dr[ew] on both theoretical and empirical foundations in the areas of ethics and assessments” (Green et al., 2007). In review of the literature, there appears to be an underlying principle that underscores ethical behavior practices being governed by judgment. Yet, the literature also highlights the importance of assessment being fair and equitable and accurately reflect student achievement levels and overall mastery of the subject matter being tested. The researcher of this study found that the general principles that govern ethics related to assessment and testing that tend to focus on the “Do No Harm” adage, fail to address what can derail that concept—an educator’s judgment or lack thereof. Thus, this study’s prevailing research questions and what guided this study were as follows:

RQ1: “To what degree do educators agree or disagree as it relates to whether or not specific assessment practices are ethical or not?”

RQ2: “Are there strong disagreement gaps between educators’ perceptions when judging whether or not an assessment practice is ethical or not?”

Strong agreement would indicate that educators’ judgments or/and perceptions about the ethical practice (of a particular assessment practice) are aligned, and strong disagreement indicates that there are gaps in ethical judgments related to the assessment practice.

Participants

Participants for this study consisted of 159 K-12 in-service teachers from an undisclosed Alabama School District. At the time of this study, the district was comprised of roughly 8,300 students, 530 teachers and 20 schools. The district consisted of four high schools, four middle schools, and 12 elementary schools. The Ethical Assessment Practice Survey was utilized to

conduct the study (Green at al., 2007). The survey was administered to 530 teachers and was completed by 159 respondents, reflecting a 30% return rate. The researcher used a 36-question scenario-based survey to question participants views (as educators) on whether or not various evaluation practices were ethical or not. Further, for the purpose of the study, the researcher broke down participants into demographics groups to include: Gender: male (N=22) and female (N=137); Degree Award: Bachelor's degree obtained (N=44) or Masters' Degree Obtained or higher (N=115); Grade Level Taught: Kindergarten-5th (N=84), 6th-8th (N=31), 9th-12th (N=44); and Years of Service: 1-7 (N=34), (N=125).

Research Design and Instrumentation

Instrument

The Ethical Assessment Practice Survey was utilized to conduct a 2007 study examining faculty perceptions regarding assessment and perceptions related to ethical practices (Green at al.). The tool was constructed and aligned to support the “guidelines for ethical student evaluation into a framework that addresse[d] both classroom assessment and standardized testing” (Green at al., 2007, p. 1002). Prior to the 2007 study, to strengthen the validity of the instrument, a piloted survey was field-tested in 2004 with 74 participants. As a result, six questions were assessed as being confusing and were either modified or replaced altogether by the researchers (Green at al., 2007). For the purposes of this research, the survey consisted of 36 scenario-based items (related to both standardized and classroom assessment practices) and was administered in the spring of 2019. Six questions were demographic, and 30 questions were scenarios that participants had to assess whether the outlined practice (described) was ethical or not. The scenarios posed and outlined in the research were developed by researchers analyzing what is acceptable assessment practices based on the prevailing and current research (Green at al., 2007). Relevancy in developing scenarios around what were commonly held ethical or unethical practices enabled the researchers to have greater confidence in the qualitative measurements of educators' perspectives, especially if there were high disagreement levels between what is or is not ethical.

The scenario-based survey instrument had seven categories of measure. They are listed as follows: Category I: Standardized Test Preparation; Category II: Standardized Test Administration; Category III: Multiple Assessment Opportunities; Category IV: Communications about Grading; Category V: Grading Practices; Category VI: Bias, and Category VII: Confidentiality. In the development of the survey instrument, each category was aligned with specific scenarios questions that addressed the overall category and ultimately was used to measure educator's perceptions regarding whether or not the scenarios posed were ethical. For the purpose of the study, participants were asked to read each scenario under the category which specifically “related to assessment issues that arise in the classroom” and respond to the scenarios using their judgment in determining if they considered the practice to be ethical or unethical (Green et al., 2007, p. 1002). The researcher determined that the same agreement and disagreement parameters used in the 2007 study, would be duplicated in this study as well (Green at al.), especially since the instrument had been field tested and vetted. Thus, 80% agreement was determined to be the threshold for agreement because it tends to show a baseline for a “high level of agreement among respondents” (Green at al., 2007, p. 1003). Of course, the higher the percentage, the higher the level of agreement. Strong agreement levels suggest “like” judgments are being made; whereas, weak judgments suggest continued discussions surrounding application

would be beneficial (Green et al., 2007, 1003). Strong disagreement was defined as items having a percentage of disagreement between 50% and 70% because it tends to show a significant level of disagreement (Green et al., 2007). Areas ranging between 70-79% were viewed as moderate disagreement. Demographic variables were strongly considered throughout the study and assessed relative to participant responses. The researcher sought to measure responses based on the following demographic variables: gender (male vs. female); level of degree obtained (Bachelor's degree vs. Master's degree); grade level taught (Kindergarten-6th, 7th-8th, 9th-12th); and years of teaching experience (1 to 7 years vs. 8 or more years).

Data Collection

This study employed a quantitative research approach for data collection, for the researcher was interested in measuring educators' perceptions about the ethics of various assessment practices. As such, the researcher determined that a quantifiable research method be used to measure participants perceptions and/or judgments regarding whether or not select assessment practices were ethical or not. This approach to collecting the data enabled the researcher to cross-analyze the data and quantifiably determine those areas of strong agreement or strong disagreement in measurable, relatable terms. Data was collected by way of a 36-item scenario based online survey (six questions were related to demographic information and 30 questions were scenario-based questions). Participant identities to include names were not requested and participation was voluntary. The survey instrument was sent district-wide and data was collected online via a survey platform. Once all survey data was collected, it was first filtered (for analysis and measurement purposes) collectively and analyzed as a whole. Then, the data was filtered based on specific demographic variables: gender (male vs. female); level of degree earned (Bachelors vs. Masters); grade level taught (Kindergarten-6th, 7th-8th, 9th-12th); and years of teaching experience (1-7 years, 8 or more years). Once filtered by the demographics data, the researcher had to manually extract and analyze the data based on the pre-set category questions that were used to measure levels of teacher agreement or disagreement (across the scenarios posed). Data from each category was collected, sorted and analyzed across all targeted scenarios to include the following: Category I: "Standardized Test Preparation", questions: 2, 4, 5, 6, 12, and 23; Category II: "Standardized Test Administration", question items 7 and 17; Category III: "Multiple Assessment Opportunities", question items 10, 21 and 25; Category IV: "Communications about Grading" contained, questions 1, 27, 29, and 31; Category V: "Grading Practices", questions 3, 8, 11, 13, 14, 15, 19, 24, 28, 30, 32, 33 and 35; Category VI: "Bias", questions 18, 22, 26, 34, and 36. And, lastly, Category VII: "Confidentiality", questions 9, 16, and 20.

Results

As noted, 159 in-service K-12 teachers across a single school district responded to the survey. Analysis is based on those findings and the demographics therein. Results from each Category (I-VII) are discussed and analyzed separately. In-service results are addressed first (under each category) and then demographic analysis follows.

Category I: Standardized Test Preparation

Agreement levels based on in-service teacher responses to the six items under the Category I: "Standardized Test Preparation" ranged from 52% to 95.6% (see Table 1). As 80% or higher is

considered to be a high level of agreement, there were four items that scored at this level. The items were Items 5, 6, 12 and 23

Disagreement levels varied. Category I identified four of the six items had demographic variables suggesting high levels of disagreement (falling within the 50% to 70% range). (Item 2) *A teacher adds vocabulary words from a standardized, norm referenced verbal aptitude test to classroom vocabulary tests* showed high levels of disagreement among females (66% thought the teacher’s practice was ethical; whereas, 34% indicated the teacher was unethical). Further, 67% of K-6 teachers indicated the teacher’s practice was ethical while 33% indicated the practice as unethical.

(Item 4) *Based on his review of the district’s mathematical framework, a teacher creates learning activities with specific math problems that are included in the annual achievement test* also showed high levels of disagreement in multiple demographic categories: 56% of teachers with a bachelor’s degree thought the practice was ethical; whereas, 44% indicated that the practice was unethical; and 52% of teachers with a master’s degree thought the practice was ethical; whereas, 48% considered the practice unethical. This trend continued with 45% of female teachers thinking the (Item 4) practice was ethical; whereas, 55% indicated it was unethical; yet 71% of males thought the practice was ethical; indicating that differing perceptions may be influenced by gender. Disagreement continued, for 46% of K-6 teachers thought the (Item 4) practice was ethical; whereas, 54% indicated it was unethical. In addition, 52% of 7-8 grade teachers thought the practice was ethical; whereas, 48% considered the practice unethical. Lastly, 64% of 9th and 12th grade teachers indicated the practice as being ethical; whereas, 36% indicated it was unethical. In assessing agreement and disagreement based on years of experience for Item 4, disagreement was still evaluated. 52% of teachers with 1 to 7 years of experience and 8 years or more of experience thought the practice was ethical; whereas, 48% considered the practice to be unethical. Items 3 and 23 also demonstrated high levels of disagreement ranging from 61% vs. 39% and 58% vs. 42% respectively.

To recap the findings, although initial findings under the Pre-service category suggested that agreement was high, when the data was analyzed demographically, strong disagreement under Category I (Item 4) was noted. Indeed 4 of the 6 items (under this specific Item number) yielded strong disagreement. And, disagreement displayed across various demographics to include degree obtainment level, gender, grade level taught and years of experience. Meaning, no demographic group that was being analyzed agreed on whether or not it was ethical for a teacher to create learning activities with specific math problems that are included on an annual achievement test. Disagreement was noted but to a smaller degree on other scenarios posed under this category. Equally important, and worth noting, under gender, a complete polar opposite response was recorded. Here the majority of males (71%) thought the instructor acted ethically, while only 45% of females did, reflecting a clear difference in attitudes and judgment.

Table 1

Percentage of in-service educators indicating the ethicality of assessment practices in standardized test preparation (across demographic descriptors)

Item#	Scenarios	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8 or more years %
	Standardized Test Prep											

2	A teacher adds vocabulary words from a standardized, norm-referenced verbal aptitude test to classroom vocabulary tests	Ethical Unethical	76.9 23.1	78.6 21.4	79 21	65.8 34.2	100 0	67 33	84 16	91 9	91 9	73 27
4	Based on his review of the district's mathematical framework, a teacher creates learning activities with specific math problems that are included in the annual achievement test.	Ethical Unethical	52.2 47.8	55.8 44.2	51.7 48.3	45 55	71 29	46.3 53.7	52 48	64 36	52 48	52 48
5	A teacher spends a class period to train his students in test-taking skills (e.g., not spending too much time on one problem, eliminating impossible answers, guessing).	Ethical Unethical	95.6 4.4	93.2 6.9	96.7 3.2	95.1 4.9	90 10	94 6	94 6	100 0	94 6	96 4
6	A teacher administers a parallel form of a norm-referenced achievement test to her students in preparation for the state testing. The parallel form is another version of the state test that assesses the same content; however, the items on the parallel form are not the same ones as on the state form of the achievement test.	Ethical Unethical	82.9 17.1	75 25	97 3	86.6 13.4	90 10	87 13	61 39	91 9	88 12	81 19
12	A teacher uses scoring high on the MAT, a commercially available publication with the same format and skills as the Metropolitan Achievement Test (but not the same items), in preparation for state testing.	Ethical Unethical	85.3 14.7	77.3 23.7	80 20	84 16	90 10	84.3 15.7	84 16	88 12	82 18	86 14
23	An elementary teacher quizzes students in the lunch line about the number of pints in a quart because students had missed the item on previous administrations of the state standardized test.	Ethical Unethical	78.7 21.3	74.4 25.6	92 8	82 18	76 24	82.5 17.5	58 42	86 14	75 25	80 20

Category II: Standardized Test Administration

Agreement levels based on in-service teacher responses to the two items under Category II “Test Administration” are as follows: There was one item that scored at a high level of agreement (80% or higher level)-(Item 7) *While administering a standardized test, a teacher notices that a child has missed a problem that the student obviously knows. The teacher stands by the child's desk, taps her finger by the incorrect problem, shakes her head, and walks on to the next desk* (see Table II): For this item, 93.4% of in-service teachers considered the practice to be unethical. On the other hand, (Item 17) *While administering a standardized test, a teacher notices that a child has skipped a problem and is now recording all his answers out of sequence on the answer form. The teacher stops at the child's desk and shows the student where to record the answer he is working on and instructs him to put the answers to each question with the same number on the answer sheet* scored at a high level of disagreement. 55% of in-service teachers indicated the practice was ethical; whereas, 45% of teachers found the practice to be unethical. Meaning consensus was not reached among survey respondents. In fact, it was nearly split evenly with some faculty asserting that the actions taken by the faculty member (within the scenario) were ethical and others scoring it as unethical.

Disagreement was also noted under Category II. In examining demographic results, one of the two items had demographic variables suggesting high levels of disagreement (falling within the 50% to 70% range). (Item 17) *While administering a standardized test, a teacher notices that a child has skipped a problem and is now recording all his answers out of sequence on the answer form. The teacher stops at the child's desk and shows the student where to record the answer he is working on and instructs him to put the answers to each question with the same number on the answer sheet* showed high levels of disagreement among in-service teachers in the following demographic areas: by degree type, gender (female), grade level taught, and years of experience.

Indeed, 44% of teachers holding a bachelor's degree indicated that they thought the practice (highlighted in Item 17) was ethical; whereas, 56% indicated the practice was unethical; while 57% of teachers holding a master's degree deemed the practice ethical; whereas, 43% deemed the practice unethical. Further, under Item 17, 52.5% of female teachers deemed the practice ethical; whereas, 47.5% deemed the practice to be unethical; and 50% of male teachers deemed the practice ethical; whereas, 50% of males deemed the practice to be unethical. Disagreement on Item 17 continued. This pattern of disagreement continued across all other demographic variables to include grade level taught, years of teaching experience (See Table 2).

Category II (Item 17) disagreement appears to cross demographic variables (as was previously noted under Item 4 in Category 1) to include degree type, gender, grades taught and years of service. And, although disagreement was noted across all variables, the highest level of disagreement fell under gender (with both males and females split on whether the instructor acted ethically or not) and by grade level assignments (K-6 and 7th -8th grade instructors) also split on the topic. Hence, again, no consensus was reached for Item 17.

Table 2

Percentage of in-service educators indicating the ethicality of assessment practices in standardized test administration (across demographic descriptors)

Item#	Scenarios	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8 or more years %
7	While administering a standardized test, a teacher notices that a child has missed a problem that the student obviously knows. The teacher stands by the child's desk, taps her finger by the incorrect problem, shakes her head, and walks on to the next desk.	Ethical	0.6	0	0	0	0	0	3	0	0	1
		Unethical	93	100	100	100	100	100	97	100	100	99
17	While administering a standardized test, a teacher notices that a child has skipped a problem and is now recording all his answers out of sequence on the answer form. The teacher stops at the child's desk and shows the student where to record the answer he is working on and instructs him to put the answers to each question with the same number on the answer sheet.	Ethical	55	44.2	57.7	52.5	50	52	52	64	42	58
		Unethical	44	55.8	43.3	47.5	50	48	48	36	56	42

Category III: Multiple Assessment Opportunities

Agreement levels based on in-service teacher responses to the three items under Category III: “Multiple Assessment Opportunities” ranged from 76% to 100% (see Table 3). As 80% or higher is considered to be a high level of agreement, there were two items that scored at this level. The two items were: (Item 21) *A teacher assesses student knowledge by using many types of assessments: multiple-choice tests, essays, projects, portfolios* (scoring at 100% agreement); and (Item 25) *A second-grade teacher uses observations as the sole method to access what students have learned* (86.6% agreement). The third item (Item 10) *A high school social studies teacher bases students' final semester grade on 2 multiple-choice tests* had an agreement level of 76%.

Disagreement was also noted under Category III. In fact, one of the three items had demographic variables suggesting high levels of disagreement (falling within the 50% to 70% range). (Item 10) *A high school social studies teacher bases students' final semester grade on 2 multiple-choice tests* showed somewhat high levels of disagreement among in-service teachers teaching in grades 9-12. Only 70% of grade 9-12 teachers deemed the practice unethical; whereas, 78% of k-6 and 77% of 7-8 grade teachers viewed the practice as unethical. Thereby, Category III (Item 10), which dealt with multiple assessment opportunities, had more agreement than non-agreement. And, even where there was disagreement, the scores were in the 70 or higher percentile. Meaning, while in disagreement, the disagreement was not as stark or split as noted among other variables tested.

Table 3

Percentage of pre-service and in-service teachers indicating the ethicality of evaluation practices using multiple assessment practices (across demographic descriptors)

Item#	Scenarios	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8 or more years %
10	A high school social studies teacher bases students' final semester grade on 2 multiple-choice tests.	Ethical	24	18.6	27.2	22.2	24	22	23	30	18	26
		Unethical	76	81.4	72.8	77.8	76	78	77	70	82	74
21	A teacher assesses student knowledge by using many types of assessments: multiple-choice tests, essays, projects, portfolios.	Ethical	100	100	100	100	100	100	100	100	10	100
		Unethical	0	0	0	0	0	0	0	0	0	0
25	A second-grade teacher uses observations as the sole method to access what students have learned	Ethical	13.4	11.9	16.3	11.25	14.3	11	3	25	6	15
		Unethical	86.6	88.1	83.7	88.75	85.7	89	97	75	94	85

Category IV: Communication about Grading

Agreement levels based on in-service teacher responses to the four items under the Category IV: “Communications about Grading” ranged from 62.9% to 99.4% (See Table 4). As 80% or higher is considered to be a high level of agreement, there were three out of four items that scored at this level. They were: (Item 1) *A teacher states how she will grade a task when she assigns it*; (Item 27) *A teacher tells students what materials are important to learn in preparing for a class test*; and (Item 31) *A middle school principal directs teachers to give students a written policy that explains how report card grades are calculated in their classes*. In contrast, (Item 29) *For the final exam, a teacher always uses a few surprise items about topics that were not on the*

study guide scored at a high level of disagreement (falling within the 50% to 70% range). In fact, 37% of in-service teachers indicated the practice was ethical, whereas, 63% of teachers found the practice to be unethical.

Disagreement, although limited, was noted under Category IV. One of the four items (Item 29) showed a high level of disagreement (falling within the 50% to 70% range). (Item 29) *For the final exam, a teacher always uses a few surprise items about topics that were not on the study guide* showed a high level of disagreement in eight of the nine demographic variables (among in-service teachers, degree type, gender, grade level taught (K-6 and 9-12), and years of experience). In review, 32% of teachers with a bachelor’s degree deemed the practice ethical compared to 68% deeming the practice unethical. In comparison, 39% of teachers holding a master’s degree found the practice to be ethical; whereas, 61% found the practice to be unethical. In addition, under the same item number, 29, 37% of female teachers indicated the practice as being ethical; whereas, 63% indicated the practice as being unethical. For males, 38% of male teachers considered the practice to be ethical; whereas, 62% of male teachers considered the practice to be unethical. Disagreement continued as it related to grade levels taught and years of experience (See Table IV). Indeed, Item 29, like similar items, had strong disagreement across all demographic variables. In each case, consensus of agreement was not reached. The greatest disagreement was found to be within the 9th and 12th grade demographic, for it was nearly split (45—ethical / 55 unethical), so again, no consensus could be reached under the scenario posed.

Table 4

Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices related to communication about grading (across demographic descriptors)

Item#	Scenarios Standardized Test Prep	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8+ years %
1	A teacher states how she will grade a task when she assigns it.	Ethical Unethical	98.7 1.3	100 0	99 1	100 0	100 0	100 0	94 6	100 0	100 0	98 2
27	A teacher tells students what materials are important to learn in preparing for a class test.	Ethical Unethical	95.6 4.4	90.7 9.3	96.7 3.3	95 5	100 0	95 5	90 10	100 0	94 6	96 4
29	For the final exam, a teacher always uses a few surprise items about topics that were not on the study guide	Ethical Unethical	37.1 62.9	31.8 68.2	39 61	36.6 63.4	38.1 61.9	36 64	29 71	45 55	40 60	44 56
31	A middle school principal directs teachers to give students a written policy that explains how report card grades are calculated in their classes.	Ethical Unethical	99.4 0.6	97.7 2.3	100 0	98.8 1.2	100 0	98 2	100 0	100 0	100 0	99 1

Category V: Grading Practices

Agreement levels based on in-service teacher responses to the thirteen items under the Category V: “Grading Practices” ranged from 72.9% to 92% (See Table 5). As 80% or higher is considered to be a high level of agreement, there were four out of thirteen items that scored at this level. The four items were (Item 3) *For a group project, a teacher bases each student's grade on the group's product and a heavily weighted individual component* which had a 92% agreement rating. (Item 14) *To minimize guessing, a teacher announces she will deduct more points for a*

wrong answer than for leaving the answer blank, which had an 82.9% approval rating; (Item 15) *To encourage lively discussion in English III, a teacher counts class participation as 30% of the final grade*, which had an 83.2% agreement rating and (Item 35) *A teacher lowers report card grades for disruptive behavior* which had a 90.3 approval rating.

Strong levels of disagreement were noted under Category V. In fact, eight of the thirteen items had demographic variables suggesting high levels of disagreement (falling within the 50% to 70% range). What is important about the results under Category V is the level of disagreement across the majority of item numbers and across the spectrum of demographic variables analyzed. Indeed of the thirteen scenarios presented to participating in-service respondents, results indicated that only 4 of the 13 items had strong agreement; and only 5 out of 13 items (scenarios posed in assessing demographic variables) resulted in a level of 80% agreement. Demographic variables included: degree type, gender, grade levels taught, and years of experience.

The highest in-service disagreements levels noted under Category V were Items 32 and 33. Item 32 dealt with the ethics of weighting homework heavily in determining final grades, and Item 33 dealt with the assigning of grades based on student growth. In-service respondents were nearly split on both items yielding a non-consensus outcome. In assessing demographic disagreements similar disagreements were noted. In fact, the strongest disagreements occurred under Items: 8, 11, 19, 32 and 33. Disagreement among these items ranged from 43% to 57% respectively, with Item 33 yielding the highest disagreement index in the category. In fact, in assessing demographic data to include: degree type, gender and grade level taught, it demonstrates that these factors attributed to the disagreement index to strengthen disagreement even further, with results for all three being 49/51 splits, indicating no agreement was reached at all. What this data demonstrates is respondents are relying on personal judgments that can be influenced by degree type, gender, grade level taught, years of experience and other factors (not evaluated under this study), opposed to common ethical principles related to evaluation being utilized to judge the scenarios. As a result, instead of strong agreement being reached (most the time), strong disagreement is being reached (50% of the time) throughout the study.

(Item 32) *A teacher weighs homework heavily in determining report card grades* showed a high level of disagreement in all nine demographic areas. Results were as follows: 38% of bachelor's degree respondents viewed the action as ethical versus 62% who viewed the action as unethical. For master's prepared respondents, 50% of respondents felt that teachers action were ethical and the other half (50%) viewed the behavior as unethical. Females and males disagreed as well where 35% classified the behavior as ethical, versus 65% unethical. Interesting, the reverse was true for male respondents where 62% of males felt the teacher's actions were ethical versus 38% deeming the actions as unethical. Regarding K-6 teachers, 35% viewed the actions as ethical, versus 65% of K-6 teachers who did not. 7th and 8th grade teachers were equally divided—with 42% deeming the behavior as ethical versus 58% deeming it unethical. High school teachers (9-12) were a little more even, but still divided, for 68% of the 9th and 12th grade teachers coded this action as ethical versus 32% who coded it as unethical. Lastly, those faculty with 1 to 7 years of experience and those with 8 or more years both indicated high levels of disagreement. 48% of teachers with less than 8 years of experience coded the behavior as ethical versus 52% who coded the behavior as unethical. For those faculty with 8 or more years, the trend continued, for 44% of respondents reported the action as ethical versus 56% who disagreed and felt the action was unethical.

(Item 33) *A teacher considers a student's growth in assigning grades* showed a high level of disagreement in seven demographic areas: for bachelor's degree, 49% of respondents stated that

the action was ethical while 51% reported it as unethical; for Master's degree, 58% of respondent coded the action as ethical versus 42% coding it as unethical. 51% of female respondents deemed the behavior as ethical versus 49% that deemed it as unethical. And, for males, 67% viewed the behavior as ethical versus 33% who viewed it as unethical. This disagreement in responses continued among K-6 teachers, for 51% of K-6 teachers found the action ethical versus 49% who viewed the action as unethical. And, among 7 to 8 grade teachers, 48% deemed the action ethical versus 52% who deemed it unethical. Lastly, teachers with 1 to 7 years of experience also disagreed, with 42% flagging the action as ethical versus 58% unethical.

Table 5

Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices related to grading practices (across demographic descriptors)

Item#	Scenarios Standardized Test Prep	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8 or more years %
3	For a group project, a teacher bases each student's grade on the group's product and a heavily weighted individual component.	Ethical	92	90.5	92.4	91.25	95.2	90	94	93	91	93
		Unethical	8	9.5	7.6	8.75	4.8	10	6	7	9	7
8	A physical education teacher gives a student a zero as a homework grade for not returning a form requiring a parent's signature.	Ethical	39.7	45.2	38.5	27.85	47.6	27	45	60	58	34
		Unethical	60.3	54.8	61.5	72.15	53.4	73	55	40	42	66
11	An accounting teacher gives a student an F for the course because the student missed the final exam.	Ethical	30	18.6	33	27.50	42.9	28	23	39	27	31
		Unethical	70	81.4	67	72.50	57.1	72	77	61	73	69
13	As a teacher finalizes grades, she changes one student's course grade from a B+ to an A because tests and papers showed the student had mastered the course objectives even though he had not completed some of his homework assignments.	Ethical	31.9	21.4	35.9	34.6	19	35	33	25	19	35
		Unethical	68.1	78.6	64.1	65.4	81	65	67	75	81	65
14	To minimize guessing, a teacher announces she will deduct more points for a wrong answer than for leaving the answer blank.	Ethical	16.8	14.6	16.5	16.5	10	16	17	18	0	21
		Unethical	83.2	85.4	83.5	83.5	90	84	83	82	100	79
15	To encourage lively discussion in English III, a teacher counts class participation as 30% of the final grade.	Ethical	82.9	81.4	83.7	87.7	81	88	71	82	88	81
		Unethical	17.1	16.6	16.3	12.3	19	12	29	18	12	19
19	A middle school history teacher offers extra credit opportunities to all his classes except the advanced class.	Ethical	26.1	16.7	34.8	16.25	28.6	17	26	43	27	26
		Unethical	73.9	83.3	65.2	83.75	71.4	83	74	57	73	74
24	A teacher lowers grades for late work by one letter grade for each day.	Ethical	60.5	53.5	67	46.25	85.7	48	61	84	64	59
		Unethical	39.5	46.5	33	53.75	14.3	52	39	16	36	41
28	A teacher uses student peer ratings as 40% of the grade on an oral report.	Ethical	21.7	19	25	18.75	28.6	18	19	30	21	21
		Unethical	78.3	81	75	81.25	71.4	82	81	70	79	79
30	A teacher considers student effort when determining grades.	Ethical	69.6	65.1	68.5	64.2	66.7	64	77	75	61	72
		Unethical	30.4	34.9	31.5	35.8	33.3	36	23	25	39	28
32	A teacher weighs homework heavily in determining report card grades.	Ethical	45.5	38.1	49.5	35.4	62	35	42	68	48	44
		Unethical	54.5	61.9	50.5	64.6	38	65	58	32	52	56

33	A teacher considers a student's growth in assigning grades.	Ethical	56.3	48.8	57.6	50.6	66.7	51	48	73	42	59
		Unethical	43.7	51.2	42.4	49.4	33.3	49	52	27	58	41
35	A teacher lowers report card grades for disruptive behavior.	Ethical	9.7	4.8	12.1	5	28.6	5	10	18	15	8
		Unethical	90.3	95.2	87.9	95	71.4	95	90	82	85	92

Category VI: Bias

Agreement levels based on in-service teacher responses to the five items under Category VI: “Bias” ranged from 73% to 95% (See Table 6). As 80% or higher is considered to be a high level of agreement, there were four items that scored at this level. The four items were (Item 34) *A teacher allows a student with a learning disability in the language arts to use a tape recorder when the student answers the essay questions on social studies tests* which reached 95% agreement that the teacher’s action was ethical; (Item 26), *A teacher always knows the identity of the student whose essay test she is grading* which reached 75% agreement that the action was ethical; (Item 18), *A teacher who knows a student had a bad week because of problems at home bumps the student's participation grade up a few points to compensate for his bad score on a quiz* 83% of in-service teachers felt the action was unethical). And lastly, for (Item 22), *Two teachers teach different sections of the same course. Because of his belief that students' work is rarely perfect, one teacher gives very few grades of "A"* (84% of in-service teachers deemed the behavior unethical).

Disagreement was noted under one item under Category VI. One of the five items (Item 36) had demographic variables suggesting high levels of disagreement (falling within the 50% to 70% range). (Item 36) *To enhance self-esteem, an elementary teacher addresses only students' strengths when writing narrative report cards* showed a high level of disagreement in all demographic areas: For bachelor’s prepared teachers, 60% found the action of the teacher to be ethical, but 40% found it to be unethical. And, for master’s prepared teachers, 56% found the action ethical versus 44% found the action unethical. Disagreement between females and males was also noted with 49% of females finding the action of the teacher to be ethical versus 51% noting it was unethical. For males, 43% of males viewed the teacher’s actions as ethical versus 51% viewing the actions as unethical. K-6 teachers were also in disagreement with 43% deeming the teacher’s actions as ethical and 57% deeming the practice as unethical. In addition, both 7th and 8th grade teachers and those teachers with more than 8 years of teaching experience had divided viewpoints related to ethical judgment; for under both categories, 48% felt the teacher’s behavior was ethical, yet 52% felt it was unethical. Disagreement was also evidenced among 9th and 12th grade teachers (57% vs. 43%) and for those instructors with 1 to 7 years of experience (47% vs. 53%) who felt the practice was ethical.

Unlike the previous Category (V) which yield high disagreement, Category VI, which required respondents to judge bias in grading, respondents strongly agreed on the ethics questions posed, opposed to disagreed. The only area of strong disagreement was noted under a single item—Item 36. Again, what is noteworthy is that in assessing demographic data, all variables (degree type, gender, grade taught, and years of experience) demonstrated strong disagreement, so consensus was not reached.

Table 6

Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices related to bias (across demographic descriptors)

Item#	Scenarios	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8 or more years %
18	A teacher who knows a student had a bad week because of problems at home bumps the student's participation grade up a few points to compensate for his bad score on a quiz.	Ethical Unethical	16.3 83.3	28.6 71.4	27.5 75.5	27 73	23.8 76.2	27 73	23 77	28 72	27 72	26 74
22	Two teachers teach different sections of the same course. Because of his belief that students' work is rarely perfect, one teacher gives very few grades of "A".	Ethical Unethical	16.2 83.8	19.5 80.5	13.3 86.7	18 82	5 95	18 82	17 83	12 88	25 75	14 86
26	A teacher always knows the identity of the student whose essay test she is grading.	Ethical Unethical	75.3 24.7	83 17	71 29	76 24	76.2 23.8	80 20	71 29	70 30	84 16	73 27
34	A teacher allows a student with a learning disability in the language arts to use a tape recorder when the student answers the essay questions on social studies tests.	Ethical Unethical	95 5	97.6 2.4	96 4	95 5	95.2 4.8	93 7	97 3	98 2	94 6	95 5
36	To enhance self-esteem, an elementary teacher addresses only students' strengths when writing narrative report cards.	Ethical Unethical	47.7 52.3	59.6 40.4	43.8 56.2	49 51	43 57	43 57	48 52	57 43	47 53	47.5 52.5

Category VII: Confidentiality in Testing

Agreement levels based on In-service teacher responses to the three items under the Category VII: "Confidentiality" ranged from 77% to 87%. As 80% or higher is considered to be a high level of agreement, there was one item that scored at this level--(Item 16). For Item 16, *A second-grade teacher uses observations as the sole method to assess what students have learned*, 13% of respondents deemed the practice ethical versus 87% deemed it unethical. For (Item 9), *A teacher adds vocabulary words from a standardized, norm-referenced verbal aptitude test to classroom vocabulary tests* it nearly met agreement, but not fully. Results indicated that 77% of respondents stated the practice was ethical versus 23% deeming it unethical, a little shy of the 80% agreement threshold.

Overall results indicate that under Category VII which looked at confidentiality in testing, there was more agreement (amongst the variables) than disagreement. However, the one item (Item 20) that examined final exam practices, had strong disagreement across a spectrum of demographic variables to include degree type, gender, grade level taught and years of experience. The one demographic group that demonstrated the highest disagreement was 9th and 12 grade instructors. For 9th through 12 grade teachers, 45% viewed the actions as ethical versus 55% unethical. And, in terms of years of experience, 39% of those instructors with 1 to 7 years of experience determined that the behavior was ethical versus 61% unethical, and for those instructors with 8 or more years of experience, 36% viewed the practice as ethical versus 64% who classified the behavior as unethical.

Table 7

Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices related to confidentiality in testing (across demographic descriptors)

Item#	Scenarios	Respondent Answers N=159	Inservice Teachers %	B.A degree %	M.A degree %	Female %	Male %	K-6 teacher %	7-8 teacher %	9-12 teacher %	1-7 years %	8+ years %
9	A teacher adds vocabulary words from a standardized, norm-referenced verbal aptitude test to classroom vocabulary tests	Ethical Unethical	77 23	79 21	79.1 20.9	73 27	100 0	67 33	84 16	91 9	91 9	73 27
20	For the final exam, a teacher always uses a few surprise items about topics that were not on the study guide	Ethical Unethical	37 63	31.8 68.1	39.1 60.9	37 63	38 62	36 64	29 71	45 55	39 61	36 64
16	A second-grade teacher uses observations as the sole method to assess what students have learned.	Ethical Unethical	13 87	11.9 88.1	16.3 83.7	13 87	14 86	11 89	3 97	25 75	6 94	15 85

Limitations

There were several notable limitations to the study. One limitation is that respondents, when responding to the various scenarios within the survey, could only select whether the behavior described in the scenario was “ethical” or “unethical”. There was no other option available to respondents, like “neither”. In addition, given this was a quantitative research study, survey respondents could not explain the basis for how they reached their decisions related to evaluating scenarios as “ethical” or “unethical”. Another limitation is that the survey instrument does not explore what type of professional development training respondents had related to the subject explored—judgment of ethical assessment practices.

Conclusion

Results of this study revealed that regardless of the level of degree type of the respondents (bachelors or masters); or the gender of the respondents (male or female); or the grade level taught of the respondent (K-6, 7-8, 9-12), or even years of experience of the respondents, there is strong disagreement among K-12 educators as it relates to identifying if specific evaluation and assessment practices (as identified in the various survey scenarios) were either “ethical” or “unethical.

Despite all other variables, judgment appears to be the key influence in decision-making. Indeed, respondents assessed 36 different scenarios (throughout this study) ranging in scope from whether or not it was ethical for “A teacher [to] add vocabulary words from a standardized, norm-referenced verbal aptitude test to classroom vocabulary tests” (Item 9) to assessing whether or not it was “either or “unethical” if “A teacher lower[ed] grades for late work by one letter grade for each day” (Item 24). In total, of the 36 scenarios posed to all participating respondents, only 18 out of 36 (50%) reached the 80% threshold indicating strong agreement on whether or not a specific practice was ethical or not. Conversely, that also indicates that there were 18 scenarios (50%) where educators disagreed or strongly disagreed (as seen in similar studies) (Green et al., 2007, Fan et al., 2019). Disagreement was seen throughout all demographic variables, without the indication of a pattern. Meaning, the disagreement appears to be random or based on sheer judgment being applied at the time of questioning. However, Category V, which dealt with grading

practices did have the highest level of disagreement with only 4 out of 13 scenarios being agreed upon. This would clearly indicate that there is not an acceptable or prevalent understanding of what is or is not consider ethical assessment practices that educators understand and apply. Rather, respondents responded and reacted based on their “gut” their own “know how”, and judgment, but if the judgment is ill-informed, it can lead to unethical assessment issues to follow.

Recommendations

Thus, it is highly recommended that post-secondary institutions and school districts (nationally and internationally) prioritize and train educators on better identifying and understanding what are ethical and unethical assessment practices. Faculty need to have a better and more rooted understanding of what ethical assessment looks like and what pitfalls they ought to avoid as it relates to the assessment and evaluation of students. Pre-service preparation and In-Service professional development should provide educators with a complete and thorough understanding of the definition of ethics in assessment and should be designed around the educators ability to improve their judgment about ethical decision-making related to evaluation of students. Educational entities, then, should be encouraged to use scenario-based training approaches to help faculty to gain a more informed and practical understanding of what constitutes ethical behavior in evaluation and assessment methods and what does not. Providing laundry list of “do’s and don’ts” is not sufficient, for a list cannot possibly cover every “day to day” situation a faculty member will face. Thus, targeting judgment and focusing on improving faculty understanding of ethical assessment practices versus unethical practices will be beneficial. In that way, the focus shifts to improving judgment and overall understanding so educators can more readily identify and apply generally accepted and recognized assessment practices that are deemed as ethical throughout the training platform. It is important, then, for educational organizations to establish ethical assessment practices that are commonly accepted and recognized within a university, college or school district.

References

- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). *Standards for teacher competence in educational assessment of students*. Retrieved April 11, 2021, from <https://buros.org/standards-teacher-competence-educational-assessment-students>
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Brookhart, S., & McMillan, J. (2019). *Classroom assessment and educational measurement (The NCME applications of educational measurement and assessment)*. Routledge.
- Creswell, J.W. (1994). *Research Design: Qualitative & Quantitative Approaches*. SAGE Publication.
- Davila, G. (2018). Current grading system detrimental to students. *The Piasano*, 57(8), 1.
- Fan, X., Johnson, R., Liu, J., Zhang, X., Liu, X., & Zhang, T. (2019). A comparative study of pre-service teachers' views on ethical issues in classroom assessment in China and the United States. *Frontiers of Education in China*, 14(2), 309–332.
- Feldman, A. (2019). Suing for bad grades? That might not be the best idea. *KYW News Radio*. Retrieved April 8, 2021, from <https://www.audacy.com/kywnewsradio/articles/feature-article/inside-law-suing-bad-grades-might-not-be-best-idea>.
- [Feldman, J. \(2018\)](#). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin.
- Fink, J. (2018). Teacher claims she was fired for violating 'No Zero' policy by not giving credit to students who didn't turn in project. *Newsweek*. Retrieved April 10, 2021, from <https://www.newsweek.com/teacher-claims-she-was-fired-violating-no-zero-policy-not-giving-students-who-1137573>.
- Green, S., Johnson, R., Kim, D.H., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23, 999–1011.
- Green, S., Johnson, R., Kim, D.H., & Pope, N. (2008). Educational leaders' perceptions about ethical practices in student evaluation. *American Journal of Education*, 29 (4), 520–530.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Falmer Press.
- Heritage, M., Harrison, C. (2019). *The power of assessment for learning: Twenty years of research and practice in UK and US classrooms*. Corwin.
- Joint Committee on Standards for Education (2003). *The student evaluation standards*. Corwin.
- Johnson, R. L., Liu, J., & Burgess, Y (2017). A model for making decisions about ethical dilemmas in student assessment. *Journal of Moral Education*, 46(2) 212–229.
- McParland, T. (2020). Lawsuit targets Columbia University for tuition refunds after COVID-19 closure. *The New York Law Journal*. Retrieved April 7, 2021, from <https://www.law.com/newyorklawjournal/2020/04/24/lawsuit-targets-columbia-university-for-tuition-refunds-after-covid-19-closure/>
- Pascal, C., & Bertram, T. (2016). The nature and purpose of assessment and evaluation within a participatory pedagogy. *Assessment and evaluation for transformation in early childhood education*. Routledge.
- Popham, W. J. (2017). *Classroom assessment: What teachers need to know*. Pearson.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Allyn & Bacon.

- Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation*, 56, 164–181. <https://doi.org/10.1016/j.stueduc.2017.12.008>
- Rhor, M. (2012). Former TSU law students take grade dispute to court. *Houston Chronicle*. Retrieved April 10, 2021, from <https://www.chron.com/news/houston-texas/article/Former-TSU-law-students-take-grade-dispute-to-3128943.php>
- Stern, J. (2017). The assessment of Ethics and the ethics of assessment. In O. Franck (Ed.), *Assessment in Ethics Education. A case of national tests in religious education*. Springer International Publishing.
- Sloan, K. (2014). Ex-student sues contracts professor over failing grade. *The National Law Journal*. Retrieved April 9, 2021, from <https://www.law.com/nationallawjournal/almID/1202660458932/ExStudent-Sues-Contracts-Professor-Over-Failing-Grade/?mcode=1202617074964&curindex=0&curpage=ALL&/&slreturn=20210318142316>.
- Tierney, R.D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation*, 43, 55-69.
- Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Pearson Education.
- Walker, T. (2016). Teachers divided over controversial ‘No Zero’ grading policy. *National Education Association*. Retrieved April 9, 2021, from <https://www.nea.org/advocating-for-change/new-from-nea/teachers-divided-over-controversial-no-zero-grading-policy>.
- Widiastuti, I.A. (2018). Teachers’ classroom assessment and grading practices (2018). SHS Web of Conferences. Retrieved April 10, 2021, from https://www.researchgate.net/publication/322760979_Teachers'_classroom_assessment_and_grading_practices
- Zaretsky, S. (2014). Law Student Sues Over Bad Grade That Destroyed His Legal Career. *Above the Law*. Retrieved April 9, 2021, from <https://abovethelaw.com/2014/06/law-student-sues-over-bad-grade-that-destroyed-his-legal-career/>.
- Zooter, F. (2019). Judicial follies: Suing the school. *The Ukiah Daily Journal*. Retrieved April 8, 2021, from <https://www.ukiahdailyjournal.com/2019/11/10/judicial-follies-suing-the-school/>.