

Imputed Meaning: An Exploration of How Teachers Interpret Grades

Author

Riley, T, Ungerleider, C

Published

2019

Journal Title

Action in Teacher Education

Version

Accepted Manuscript (AM)

DOI

[10.1080/01626620.2019.1574246](https://doi.org/10.1080/01626620.2019.1574246)

Downloaded from

<http://hdl.handle.net/10072/387650>

Griffith Research Online

<https://research-repository.griffith.edu.au>

Imputed Meaning: An Exploration of How Teachers Interpret Grades

Abstract

Instruction, assessment, grading and reporting are the métier of the profession, teachers find it challenging and often burdensome to provide justification for the decisions they make. This article explores teachers' thinking about the grades assigned by their peers and their explanations of discrepancies between their interpretations of the same grades assigned to different students based upon a think aloud task in which 21 teachers placed twenty-four fictional student record cards in remedial, average or advanced programs based upon student grades earned from grades four through seven. Teachers' responses to the task revealed the many meanings teachers impute to grades and the opportunities for biases to intrude in the interpretation process. This study provides concrete evidence of how subconscious biases may influence the decisions teachers make about their students. Recognition of the possibility that their decisions may be biased is a first step in modifying the ways they think about the meaning they assign to grades.

Keywords: achievement measures; assessment; bias; grading practices; teacher education

Imputed Meaning: An Exploration of How Teachers Interpret Grades

Evaluating student work is a hazardous process. Among the perils are vague criteria, inconsistent application of criteria over time and events, and bias (Bennett, Gottesman, Rocks, & Cerullo, 1993). Student behavior and the social and emotional demands that students make can easily influence their teachers' judgment unless teachers conscientiously separate their reactions from their judgments. Brookhart (1994) observed that, despite variation in methodological rigor among studies devoted to the topic, findings were remarkably consistent across geography, schools and methods, namely that teachers make significant attempts to treat students fairly. Although achievement measures such as tests are the major component in the assignment of grades, effort and ability are also commonly considered. Elementary teachers employ more informal evidence and casual observation than secondary teachers, who tend to use tests and other formal written activities. Teachers perceive the purpose and meaning of grades differently, and their perceptions are influenced by the importance they assign to achievement and non-achievement factors. It is the latter observation that has led some to use the term hodgepodge to describe teacher-grading practices (Cross & Frary, 2009; McMillan, Myran, & Workman, 2002).

Cognizant of the challenges that attend the process of assessing student work and expressing appraisals economically and precisely to other audiences, many jurisdictions spend considerable resources providing professional development and guidance. This includes educating teachers about assessment and the translation of assessments into grades, which are recorded in permanent record cards and reports to parents and guardians. In 2010, the Ontario (Canada) Ministry of Education published a 159-page guide to assessment, evaluation and reporting. Ontario provides teachers with performance standards for achievement to avoid the unfairness to students of variability among teachers in the assessment and evaluation of student

work (see Table 1 for performance standards for achievement in Grades 1 to 12). Teachers are encouraged to "assess and evaluate student work with reference to established criteria for four levels of achievement that are standard across the province" (Ontario Ministry of Education, 2010, p. 19). In addition to performance standards and practical information about assessment, evaluation, and reporting, the guide suggests that, "to the extent possible, the evaluation of the learning skills and work habits, apart from any that may be included as part of a curriculum expectation, should not be considered in the determination of letter grades or percentage marks for subjects/courses" [emphasis supplied] (Ontario Ministry of Education, 2010, p. 55).

Notwithstanding the significant efforts to improve assessment practice and the expression of assessments in report cards, there are teachers who, though they claim to grade solely to report academic achievement, include sources of evidence that are not indicative of achievement (Imperial, 2011), including the physical attractiveness of the students (Dunkake, Kiechle, Klein, & Rosar, 2012). The differential perceptions that teachers have of the meaning and purposes of grades raise questions about claims that teachers' attempt to treat students fairly. As one upper elementary teacher in Alberta, Canada candidly put it:

I know things are happening that should not be happening. The grading, the recording, it is not based on what [the students] know but on how they play the game, and I question whether the assessments are actually assessing the curriculum outcomes . . . and it is also a power thing [with the teacher who says], 'If you don't listen to me then. . . .' (Scott, Webber, Aitken, & Lupart, 2011, p.105)

Interest in teachers' interpretations of the grades assigned by their peers is derivative of a larger concern about teachers' biased judgments. Although education and assessment practices have changed over time to take into account "assessment for learning" as well as "assessment of

learning" (Bennett, 2011), assessments are judgments. The judgment process and the imputed meaning that teachers assign to the grades assigned by their predecessors are the focus of this study.

The literature devoted to the topic (Author 1 & 2, 2008 & 2012; Jussim & Harber, 2005; Bowers, 2009; Mertler, 2004; Smith, 2003; Ziomek, & Svec, 1997) makes clear that the assessment of student work and the assignment of marks to that work are problematic. The problem is manifest in vague assessment criteria, inconsistent application of criteria, and teacher bias. This paper examines the meanings that teachers assign to the grades assigned by other teachers in the context of making decisions about the placement of students based on the peer-assigned grades. The examination, informed by attribution theory informed by theories on expectations and stereotypes, intended to develop a better understanding of how teachers interpret grades and is an extension of the work of McMillan (2003) to the potentially 'high stakes' decisions teachers make when making recommendations about the opportunities afforded students (Sobkin & Kalashnikova, 2012).

Citing the author's previous work investigating decisions that teachers made about assessment and grading (McMillan & Nash, 2000), McMillan (2003) posited that the tension between their internal beliefs and the external constraints placed upon them manifested itself in the difficulty that teachers exhibited in providing coherent rationales for the decisions they made. He drew particular attention to the idiosyncratic and highly individualized nature of teacher decision making about assessment and grading describing them as "a hodgepodge of factors and influences" (p. 38). According to McMillan what is important in the teachers' decision-making process is the "effect of the decisions on student learning and motivation" (p. 38). Principles concerning "scientific measurement" such as validity and reliability "may be irrelevant to the

pragmatic needs of teachers" (p.38). Although the focus on their pragmatic needs and "the more general goals they have for student learning" (p.38), fairness (avoiding biased decisions) and the consequences of unfairness do not appear to be factors consciously considered by teachers.

The literature prompted us to re-analyze data we had collected because we believed that it would shed light on the issues discussed in the literature and extend McMillan's contribution to teacher decision making. By re-evaluating our existing study through this lens, we can provide additional insight into the basis of teachers' decision making thus highlighting important areas worthy of deliberation in the on-going discussions about grading.

Teachers' perceptions of grading

When used appropriately, grades enable teachers to communicate how well a student may understand a particular issue or concept. Grades can communicate a student's conceptual and procedural knowledge, locate a student in relation to peers, and indicate where lessons may need to be changed or adapted for individuals (Randall & Engelhard, 2010). Grades also influence the educational opportunities a student receives, and are often used to determine classroom placements, academic scholarships, and university and college admission, all of which have potential long-term influence over student's overall educational, socio-economic, and health outcomes (Alm & Colnerud, 2015; Fischbach et al., 2013; Keller, 2016). An evaluation of a student as reflected in a grade may influence a student's self-esteem (Hoge, Hanson, & Smit, 1990), as well as the effort they put towards educational endeavors in the future (Keller, 2016). Yet grading is regarded as one of the most challenging aspects of teachers' work (Linn & Miller, 2005) and continues to be described as a source of concern and anxiety amongst teachers (Thorndike, 2005). Some researchers (Alm & Colnerud, 2015; Brookhart et al., 2016) argue that

grading is often an area overlooked in initial teacher preparation programs, and as a result, new teachers are often poorly prepared for grading. Other researchers note that even with grading guidelines, there is little consistency in teachers' grading across schools (McMillan, 2008; Sun & Cheng, 2013).

Various internal and external factors influence teachers' grading practices, for example, a student's perceived effort, behavior, work habits (Guskey, 2011; Yesbeck, 2011), class participation (Cross & Frary, 1999), teaching grade level, assessment training received, class size (Cheng & Sun, 2015) and even a teachers' disposition while marking. A teacher's learning and teaching philosophy may also influence a student's grades. A study by DeBoer, Anderson, and Abdulaziz (2007) found that, while some instructors take pride in assigning few 'As', other instructors regard the notion of a student having difficulty understanding a concept in their classroom as a reflection of their ability. These two outlooks lead the researchers to hypothesize that when it comes to grading variance, "teachers' attitudes of ascription of responsibility and need for social approval play important roles" (p.61).

Svennberg, Meckbach, and Redelius's (2014) study of physical education teachers found that teachers' inability to verbally express grading criteria results in dependency upon "gut feelings" as a way of assessing students, resulting in problems of validity and the reliability of the grades they assign. Interviews revealed that teachers regarded grades as tools to reinforce student behavior, assist in classroom situations, develop student self-confidence, and encourage interaction with others. Brackett, Floman, Ashton-James, Cherkasskiy and Salovey (2013) found that teachers' moods also influence grading standards. Teachers who focus on a negative memory while grading a student would judge that student as less capable than a teacher focusing on a

positive memory, demonstrating that even a teachers' state of mind could influence grading practices.

Some researchers (Bishop, 1992; Wilson, 1996) describe the ethical dilemma some teachers experience when faced with having to give a bad mark to a student who put in the effort, but still did poorly (Colnerud, 1997; Pope, Green, Johnson, & Mitchell, 2009). The division teachers feel is between being a judge or "objective observer", focusing exclusively on achievement outcomes, versus that of a coach whose role is to encourage, praising effort and improvement. Bishop (1992) suggests that these roles are incompatible and argues that teachers should focus more on their mentoring role, leaving their judging role to external assessors. Brookhart (1991; 1993) uses Messick's (1989) validity framework to better understand the differences that exist between teachers' recommended and actual grading practices and identified that teachers' concern over the uses and potential consequences of grades (that is, teachers' awareness of how a grade could be used to either help or hinder a student) could influence grading outcomes. Bonesronning (1999) argues that grading practices may be so entwined in teachers' characteristics that grading practices cannot be changed unless teachers' characteristics change, and that "only the laudable and most experienced teachers know the true relationship between grading and real achievement" (p.103).

McMillan and Nash (2000) note that most teachers are aware and often acknowledge non-achievement and achievement factors when considering grades. However, Brookhart (2004) cautions that while teachers have the liberty to assess non-achievement factors such as effort, participation, attitude and behavior, these factors should be considered outside of grading. When these factors get conflated with achievement, the meaning of the grade becomes confused. Randall and Englehard (2010) find that non-achievement factors are more likely to be taken into

consideration by teachers when students are at grade "borderlines", meaning that teachers are more likely to consider a student's behavior or effort in situations where a student is sitting on the cusp of one grade or another. They observe that "a student with excellent behavior and low achievement is far more likely to pass a course than a student with inappropriate behavior and low achievement" (p.1379). Despite this, they state that teachers generally base "decisions on their personal beliefs and expectations which tend to support and promote student success" (p.1378). However, problems may still arise if a teacher's perception of a student's effort or behavior does not match the reality of that student's experience.

Teachers' Judgments

Ready and Wright (2011) describe teachers' judgments as accurate, inaccurate, or biased. Accuracy is desirable in assessment because it is presumed the teacher is making a judgment about performance as opposed to other attributes. When teachers can successfully predict a student's achievement without influencing it, their judgments are regarded as accurate (Jussim, 2012) and defined with the student's previous performance (Hinnant, O'Brien, & Ghazarian, 2009). When teachers' perceptions of students based on valid predictors of achievement differ randomly, they are considered to be inaccurate. However, if teachers' perceptions differ systematically, they are considered to be both inaccurate and biased (Mizala, Martinez, & Martinez, 2015; Ready & Wright, 2011).

Perceptual bias occurs when teachers base evaluations on their perceptions of students' abilities or behavior as opposed to their demonstrable achievements (Jussim, 2012; Kolb & Jussim, 1994). Teachers' perceptions of effort and concerns for students' moral development could also influence teachers' grading practices (Zoeckler, 2007). Fischbach et al. (2013) find that teachers' judgments of intelligence reflect a student's grade point average (GPA) rather than

the student's actual IQ. Since a student's GPA is still a summary of teachers' judgments, their study suggests that perceptions of student intelligence are biased towards teachers' perceptions of that student. Within the context of assessing students, teachers' may attribute a student's ability to various factors, including a student's race, ethnicity, gender, SES status, ability, effort, social skill, life circumstances and support (Sudkamp, Kaiser, & Moller, 2012). Sun and Cheng (2013) note that "the grading decisions teachers make, convey their values, beliefs, and assumptions about teaching and learning, which are rooted in and shaped by the sociocultural and historical contexts where instruction takes place" (p.327). Teachers may perceive behavior or effort based on their experiences or belief systems.

Attribution Theory, Expectations, and Stereotypes

Attribution theory is concerned with how an individual may interpret events and how this interpretation may relate to personal values. According to attribution theory (Weiner 1974; 1985), individuals identify specific causes to explain why a particular event or outcome occurred. Within the context of assessing students, internal causal attributions such as ability and effort are commonly applied, with "ability" being regarded as stable and uncontrollable, and "effort" seen as unstable and under the control of the student (Weiner, 1974; 1985). The causal attributions a teacher makes about a student may influence the teacher's affective and emotional responses to that student (Keller, 2016).

Tollefson and Chen (1988) contend that teachers who attribute a student's lack of understanding of something within the student's control are more likely to feel frustration towards the student. Teachers who attribute a student's lack of understanding to low ability (internal; uncontrollable) were more sympathetic toward the student since low ability was

perceived as being outside of the student's control. Matteucci and Gosling (2004) find that teachers' attributions of a student's behavior could influence their grading practices towards that student. Teachers who believe students did not prepare for an exam may be more likely to punish the student with a lower grade, while students perceived as trying hard were more likely to be rewarded since "effort" is regarded within a student's control. A teacher may be more lenient in her grading if she perceives an unsatisfactory grade was influenced by factors beyond a student's control (DeBoer, Anderson, & Abdulaziz, 2007).

Some research on assessment and grading practices indicates that the way teachers perceive these attributes and the degree to which they influence teachers' judgments will also be influenced by their values and beliefs. Teachers who value effort, and equate effort with time, may be inclined to highly rate students perceived to have spent more time on an assessment task (Keller, 2016). Teachers also tend to have higher expectations of students they perceive as being assertive, independent and self-confident (Rubie-Davies, 2010) and lower expectations of students they perceive as misbehaving (Clark, Thompson & Vialle, 2008) or as being disinterested (Legewi & DiPrete, 2012). They may also be more sympathetic towards learners they perceive lack ability, as opposed to those who are perceived as unwilling to put in the effort (Georgiou, Christou, Stavrinides, & Panourea, 2002; Matteucci & Gosling, 2004; Matteucci, 2007). McMillan and Nash (2000) find that some teachers use effort as a way to reward students, "giving a student who tries harder the higher grade" (p. 31). They caution that while this may help to motivate students since effort is deemed controllable, it may lead to a belief amongst students that high grades should reflect the effort they put in rather than overall mastery of a concept.

Timmermans, de Boer, and van der Werf (2016) note that, "different teachers' expectations, or teacher expectation bias, can occur when teachers systematically expect too much or too little from specific groups of students" (p.3). Earlier studies show that teachers' expectations may vary for different groups of students. Recently, more researchers (Author 1 & Author 2, 2012; Author 1, 2014a; 2014b; Reyna, 2000; Sarra, 2011) have considered the influence teachers' perceptions and stereotypes regarding certain groups may have upon their educational outcomes. By using attribution theory, in tandem with theories on expectations and stereotypes, one can begin to identify how teachers' expectations of students may be influenced by specific external and internal characteristics or traits that are consistently associated with particular sub-groups of students (Author 1, 2014a; 2014b; Reyna, 2000).

Stereotyping describes the overgeneralization of what a particular group is or does without taking into account the individual differences within that group. While some stereotypes may prove to be accurate (Jussim, 2012), when applied without evidence or despite evidence to the contrary, they could lead to biased judgments. Some teachers may associate attributes such as "low effort" or "well behaved" to specific groups of students. A teacher who then values "effort" but perceives male learners as having a lower work ethic than their female counterparts may be more likely to place male students in lower level classrooms. Studies demonstrate that teachers generally perceive female students as better behaved and exerting more effort than their male counterparts (Auwarter & Auguete, 2008; McCormick, 1994; Siegle & Reis, 1998; Timmermans et al., 2015) and that this can sometimes lead to biased decisions detrimental to both male and female learners (Author 1, 2014b; Bianco, Harris, Garrison-Wade, & Leech, 2011).

Other studies (Kelly & Carbonaro, 2012; Shepherd, 2011; Timmermans et al., 2015) demonstrate that students from minority and lower-socio-economic backgrounds may be more

susceptible to teachers' negative perceptions. In addition, the stereotypes a teacher holds, either consciously or unconsciously, may be directly or indirectly communicated through the teacher's behavior. Van den Bergh, Denessen, Hornstra, Voeten, and Holland (2010) find that teachers who are identified as having negative prejudicial attitudes towards ethnic minority students are more inclined to evaluate these students as having less academic potential. Timmermans et al. (2015) reveal that teachers' different expectations of boys and girls could be partially explained by teachers' perceptions of students' work habits, with teachers perceiving girls as having better work habits. The messages communicated to students because of teachers' erroneous perceptions may influence the way a student is treated and evaluated and may influence a student's self-perception regarding her ability to achieve (Steele & Aronson, 1995).

Teacher Decision Making Task

The review of the literature reveals the disparate nature of the grading process and the application of standards. While much of the research literature on teachers' grading practices is based on quantitative data (Brookhart, 1993, Duncan & Noonan, 2007), less work explores how teachers think about assessment and grading practices and the justifications they offer regarding the decisions or recommendations they make. This prompted us to probe the judgments that teachers made of the grades assigned on fictional student record cards and the meaning they imputed to those grades. The authors conducted an initial study (Authors, 2012) in order to determine the extent to which external factors such as a student's race, gender, or ESL status would have upon teachers' decision making.

A sample of twenty-one teachers (eighteen female and four male) with at least two years teaching experience were recruited from a metropolitan area in Western Canada to take part in a

study where they would review and place 24 fictional student record cards into one of three folders (remedial, regular, advanced). We were primarily interested in interviewing teachers with experience teaching grades six, seven, or eight, given that decisions made regarding students within these years are more likely to determine placement in secondary school. Eighteen of the teachers interviewed taught in elementary school, and three taught grade eight in secondary school. Eight teachers noted having taught various grades (Table 2). Eighteen of the twenty-one teachers interviewed had over five years of teaching experience. All teachers interviewed were of Euro-Canadian background and one teacher identified as bilingual (English/French). This sample is generally representative of the Canadian teaching body, which indicates that 72.6% are female (Canadian Teachers' Federation, 2012) and of European-Canadian background (Ryan, Pollock, & Antonelli, 2009).

Data Sources and Data Collection

Since the objective of the task was to demonstrate how small cues like a student's name or ethnicity might instigate biased perceptions or placement recommendations, the study needed to be based on a concrete, measurable task. Teachers were invited to assess twenty-four fictional student record cards, which included a limited number of variables that could influence their decisions (see Figure 1). The fictional record cards (Figure 1) were minimally designed to look specifically like those the teachers typically encounter. Each record card contained the student's grades and just enough information to elicit some form of discussion from teachers regarding the rationale for their selections. All cards included the following information; a) student's name (first, last), gender (M/F), year of birth, b) whether or not the school board had received targeted funding for one or more years based upon the student's group (Aboriginal, ESL, Non-Aboriginal)

status, and c) academic history as represented by letter grades alone from grades four to seven in six subjects (Language Arts, Mathematics, Science, Social Studies, Music, Art).

The record cards were divided so that eight records cards represented Aboriginal students, eight represented ESL students, and eight represented non-Aboriginal and non-ESL students. The fictional students' prior achievements were manipulated to ensure that students in each of the three categories would have same grades across four levels of achievement: low (6 students), low-medium (6 students), high-medium (6 students), high (six students). If a teacher made their decisions based on a student's prior academic achievement alone, their placement would consist of the following: low (six low achieving students); regular (six low-medium and six high-medium level students) and high (six high-achieving students) (see Table 3). The letter grades available on the cards were systematically varied within each category of students to eliminate recognition or bias attributable to sequencing.

When making the fictional student record cards, letter grades were selected as a "relatively unambiguous marker of student achievement" (Author 1 & 2, 2012, p.309), yet teachers' qualitative responses to the record card stimulus revealed the extent to which teachers' imputed meaning to those letter grades despite the limited information available. Teacher interviews were conducted by one of the researchers with teachers at their school during a time of their choice and were audio-recorded with the teacher's permission. Prior to the interview, teachers were reminded that the purpose of the study was to explore the nature of teacher judgment. Each teacher was provided a brief explanation of the interview task, both orally and in writing. Each teacher was asked to review the fictional student record cards and individually place them into one of three folders (supplementary learning assistance, regular grade eight

program, accelerated advanced program) based on the grades they had earned (emphasis supplied).

In order to gain a more nuanced understanding of the thought processes involved when teachers make decisions regarding student placement, teachers were invited to explain their rationale for decisions using a technique described as the "think aloud" method. This method (Van Someren, Barnard, & Sandberg, 1994) stems from psychological research and involves presenting the participant with a specific task and asking participants to articulate their thought process while performing that task, to enable a more thorough understanding of their thought processes as they complete an assigned task. The "think aloud" technique is thought to work best with people who have higher-level verbalization skills and have some level of expertise in relation to the task. Teachers, with their high level of verbalization skills and knowledge of the teaching profession, make ideal candidates as relative "experts" in their field, in that they each had completed a requisite teacher education program. However, the task is still difficult enough to ensure that the risk of teachers giving an automatic response is minimal.

Throughout the duration of the task, teachers were encouraged to "think aloud" while making their decisions. They were also invited to follow up on particular comments they made about cards during the "think aloud" experiment. For example, if a teacher noted surprise, confusion, or disappointment while reflecting upon a student's record card, they were asked to clarify their reaction, enabling the researcher to gain greater insight into their choices. All verbal responses were recorded and saved for later analysis to see if significant patterns emerged. Decisions made by teachers where record cards did not align with the selected folder (i.e., a record card with high grades was placed in a remedial folder) were identified and set aside by the researcher for further questioning.

Once the task was completed, teachers were asked to reflect upon the comments and the decisions they made throughout the process. For example, if a teacher decided to place two students with identical grades into different folders, the teacher was invited to look at the two cards simultaneously to a) note if they could identify any difference between the two cards; b) speculate upon the reason why they placed the two cards into different folders. When teachers' judgments coincided with the grades on the record card (for example, a high achieving student was placed into an advanced level classroom), that decision was deemed as accurate, meaning that the placement decision aligned with the grades a student had earned. A teacher whose judgment was influenced by factors outside of the grades achieved on the record cards (such as gender, ethnicity, or ESL status) was described as exhibiting bias. Finally, teachers were informed that the study was intended to examine the extent to which teachers' decisions about student placement may be influenced by a student's ascribed characteristics. Teachers were then given the opportunity to reflect upon their decisions or withdraw from the study. Significantly, no teacher opted to withdraw from the study.

Analysis

Interview transcripts were sent back to teachers for final review. Approved transcripts were read and re-read for themes as well as individual concepts relating to the literature review for the initial study. A strict guideline was created using a method adapted from Boyatziz (1998) and outlined by Rubin and Rubin (2005, p.216) to ensure the relevance of each concept and its legitimacy in relation to the findings. The literature related to teachers' perceptions (e.g. McMillan, 2003) and attribution theory were used to identifying three overarching themes (decision-making rationale, teachers' perceptions, beliefs, and values; external attributes). Sub-

categories were identified (e.g., grades as codes; anecdotal comments; subject matter; student behavior; EAL status, Indigenous status, gender) and paired with overarching themes. Responses sharing the same category were re-examined to confirm particular relevance to the category applied, and as a group, labeled within that category. Table 4, adapted from Conostas's (1992) "Documentational table for the development of categories" (p.262) provides further illustration of what specific elements were considered for category development. Table 5 details examples of teacher statements and the associated categories that were used. Once categories were identified, they were coded and processed using Atlas software. This article focuses primarily on the emergent themes identified in relation to teachers' responses to grades.

Findings

Findings overview

The primary objective of this study was to identify the extent to which teachers' judgments were influenced by external factors such as a student's race, gender or ESL status. Study findings revealed that of the 21 teachers interviewed, only one teacher placed student record cards accurately according to three levels of achievement, five teachers placed all students into two levels of achievement (with no cards placed in remedial classrooms), four teachers refused to differentiate between the cards despite the clear differences indicated within the students grades (meaning six high achievers were denied the opportunity to go into advanced classrooms), and eleven teachers demonstrated bias towards one or more students on the basis of factors such as race (Author 1 & 2, 2008; 2012), gender (Author 1, 2014), and EAL status (Author 1, 2014). This means that while the cards were all identical across four levels of achievement, different students were provided different opportunities not only on the basis of

their actual achievement as represented by the grades but also because of the teacher's values and beliefs.

The "think aloud" process revealed that, of the twenty-one teachers, fifteen teachers expressed apprehension about making recommendations about student placement based solely on the grades the students had earned. Teachers reasoning about program placement on the strength of grades revealed their reservations about the grading process as well as the significance of their personal beliefs and values on their decision-making. Based on teachers' comments, student placement decision-making was conceptualized to represent a fraught relationship among the following three themes (Decision Making Rational; Teachers' Perceptions, Beliefs and Values; External Attributes). This tension was revealed through teachers' responses and justifications for their decision-making. Following is a discussion on how the aforementioned findings help demonstrate the potential for teachers' unchecked perceptions to disproportionately negatively influence students. Pseudonyms have been applied throughout.

Decision Making Rational: Grades as Codes

As was the case in McMillan's (2003) study, one of the most common findings was the difficulty with which teachers had in explaining the rationale behind the decisions they made. Without direct contact with the student, grades were seen as codes that teachers found difficult to decipher and, thus, made reference to a "hodgepodge of factors" (Brookhart, 1991; Cizek et al, 1996; McMillan, 2001) that had little to do with students' prior academic performance. Some teachers based their decisions primarily on their individual perceptions and classroom experiences. Fifteen teachers expressed anxiety about making decisions based on grades alone. One teacher notes,

I know how arbitrary they [grades] are. I know the kinds of work that will determine, you know, whether a student turns out to be an A, B+ or a C+ in any term. Like the range can be quite, quite different and I respond to that.

These teachers also argued that grades don't adequately articulate the way a student would interact in the classroom and so while they were able to acknowledge high achieving students' grades on the record card, they expressed apprehension about placing high achieving students into academic classes without knowing how the student would behave in the classroom. Sarah explains,

"I wouldn't be comfortable [making placement decisions] before March so that I really understand a student. So I would be looking at how independent they are, how motivated they are, their social development".

Some teachers expressed that, while a letter grade might indicate a student's achievement, it would not provide insight into the student's potential and so, without having had any personal experience with the student, they found it difficult to make decisions without having access to former teachers' anecdotal comments. Sharon observed, "It's hard when there are no comments, and it's just letter grades because I automatically go for the comments as well." One of the reasons teachers wanted to read the anecdotal comments was to gain insight into the perspectives and practices of former teachers. For these teachers, how much credence the grades would be given would at least partially depend upon their interpretation of the former teachers' practices and philosophy. For example, some questioned the legitimacy of the record card grade because they doubted the reliability of the former teachers' judgment. Rebecca noted that "C-, a lot of the times they [the students] need it and sometimes it's "teacher code" for 'you've really failed, but I

don't want to give you an F, so you get the lowest grade I can give you". According to Rebecca, certain grades may have alternative meanings that can only be deciphered by other teachers with the assistance of anecdotal comments.

The value judgment of a previous teacher was another factor that some teachers considered prior to placing a student. Whether or not the previous teacher's values were perceived as coinciding with the participant's values regarding creativity, originality, outcomes, and effort influenced how the teacher interpreted the grades assigned to learners, as Emma's response emphasized:

What kinds of skills was the teacher looking at when that mark was given? Is it from the same teacher? Is one teacher looking at effort and the other one only looking at a finished product? Is one teacher taking into consideration creativity and originality and maybe writing and the other one maybe not as much? So I guess, to judge a kid and whether or not they belong with anyone else is really difficult for me when all I'm looking at is a letter grade because I have no information about where that letter grade came from.

Because of this, some teachers had trouble making decisions about the student record cards of ESL learners since they could not determine whether the grade represented the student's "true" academic ability or was the result of a teacher who used high grades as a form of encouragement.

Rebecca explained:

I find that that happens sometimes. It's like, you know, 'well, this is an A. It would be an A if they weren't ESL, so we'll give them an A', so people give them an A and it's like, NO! It is language arts in English! You need to mark it based on the same scale as everybody else.

For Sharon, knowing who did the grading was important because she believed a student's grade would vary in accordance with the teacher's selected teaching methods. When looking at a student's science grade, she considered, "Science tends to be a subject area where some students will excel if it's very hands on and different teachers have different sorts of styles of teaching science. Some it's more text, some it's more hands on". Sharon stated that knowing the way the course was delivered would help her decipher any discrepancies in grades as well as determine student placement. For example, although she possessed no information upon which to make the claim, she attributed fictional student's "Stanley Kyun-Yin" sudden drop in grade seven science to a "teacher who used more text material" whereas the teacher in "grade four, five, six was more hands on." Given this assumption, she believed that with the right "hands on" approach to teaching, Stanley [a low-level student] would be able to achieve in a regular classroom. Teachers' reluctance to place students based on academic grades alone highlights the importance that many of the teachers interviewed attached to factors external to students' reported achievement. This is a crucial point to consider when assessing the way in which teachers' beliefs and values may influence a student's educational achievement.

Teachers' Perceptions, Beliefs and Values

In their research on the sociology of the school grade, Sobkin & Kalashnikova (2012) note that, for some teachers, "a grade can serve as a tool for control over social behavior or act as a motivator that activates the student's cognitive activity and encourages him [sic] to pursue social achievement" (p.74). In this study, teachers acknowledged finding it difficult to make decisions based on grades alone because of their perception that grades can be used to reward or punish behavior. Without access to information about the teachers making the assessment, participants questioned the legitimacy of the grades on the card. Andrea explained:

I've worked in twelve different schools, and I know how various teachers work. I know that marks can be highly punitive. I know that they can be a way of rewarding behaviors or not. So the idea of making judgments about the suitability of programs just simply based on grades makes me very anxious. So all of a sudden, I'm resisting this as an activity.

While Andrea's comments may appear to be overly judgmental in relation to her colleagues' noble intentions, comments made by her and by other teachers revealed that her perception might not be far from the mark. One teacher belief, also reported in other research (Randall & Englehard, 2010; Stanley & Baines, 2001, Sun & Cheng, 2014) was that students' should be rewarded for behavior deemed positive within the educational system. McMillan (2003) refers to this as "pulling" for students. In other words, due to teacher's desire for their students to succeed, they may make certain allowances ("accepting late work or revisions") for students or reward students for behavior outside of academic achievement such as "giving extra credit, and including ways to earn points for trying" (p.36) for behavior they would like to reinforce. Sun & Cheng (2014) state that, "teachers tend to consider these non-achievement factors in grading because they are traits that teachers cultivate and regard as important for students' ultimate achievement" (p.328). Students' perceived by teachers as trying harder are given higher grades than students who were perceived as lacking motivation or a work ethic.

In this study, several teachers acknowledged grading students for their perceived effort, leadership, and maturity. However, teachers also revealed that if they perceived the student as expressing anxiety about achievement, they would be less likely to recommend the students for advanced placement even if their grades warranted it. Andrea explains,

Sometimes kids have really strong work ethics but their stress level with that work ethic is really intense, and they may not be able to handle it even though they've got all these As. They may not be able to handle the intensity of this advance program where you're going to lose their electives so I wouldn't recommend it.

Andrea and the other teachers consistently mentioned behavioral factors such as effort, work ethic, work habits, and a student's general attitude as factors that influenced their grading decisions. Some teachers seemed more willing to reward a student if they perceived them to be socially mature and a hard worker. Tanya reflects:

Comments are really important for me on a report card in terms of looking at how well a child's work habits are, or how strong a child's work habits are because they'd have to have really strong work habits to do well in a program that they have to be on top of things constantly.

Some teachers acknowledged that the perceived neatness of a student's work could also influence their assessment grades. One teacher discussed how the marks of grade seven male students improved when the school purchased laptops since it made their assessments neater. Andrea notes:

One of the things boys really struggle with is printing neatly. They're just not as neat and tidy as girls. It's not as important to them. Their hand-eye coordination for small tasks like that is just not as good. We have laptop programs for grade sevens [...], and I've seen

the boys doing better with their assignments because it spell-checks. It's neat. Because of that, they get a better mark right from the beginning.

A number of teachers expressed apprehension about making decisions about students they did not know because they believed the record cards did not convey sufficient information to gauge a student's potential. Steve, a teacher for nine years, agreed:

I guess I'd have to say that, not knowing the individual students, this is somewhat challenging to base this solely on a report card because a report card grade is pretty limited in its scope as to what a student is capable of achieving, producing and doing.

Steve went on to explain that "big picture" students may not necessarily be high achievers in written tests and so he would like to know whether student record grades are more reflective of a student's performance in a written test, or of the teachers' overall observations of the student's performance, and rewarding them on that basis. Olivia also acknowledged using grades to "reward" students for what she perceived as the extra effort they devoted to particular areas.

Randall & Englehard (2010) caution that, "when final grades are composed of some combination of achievement, ability, behavior, and effort, problems may arise over the meaning of the grade," (p.1372). They go on to describe how effort could be perceived in a myriad of ways ranging from attendance to participation in class and caution that low-achieving students may be more likely to benefit from "effort consideration" (p.1372) than their high-achieving peers, due to teachers desire to help students succeed ("pulling" for the student). Protiz (2016) notes that even in subjects such as math, where teachers' reported having a more standardized system of grading, teachers acknowledged taking participation into account for weaker students.

Another important influence on teachers' placement decisions were teachers' perceptions about the relative importance of different subjects. This is not surprising given the increasingly high expectations for academic achievement and institutional accountability. Proitz (2016) states that subjects in arts and physical education are more likely to be regarded as supportive subjects and compared unfavorably to courses regarded as "more important" (p.559). Proitz notes that "teachers' grading practices may also reflect varying conceptions about the function and status of subjects in the national curriculum relative to one another" (p.559).

Of the twenty-one teachers interviewed, sixteen acknowledged that it was the student's achievement in "core courses" such as math or language arts that would have the greatest influence over their placements decisions, followed by a student's consistency in maintaining high grades. For example, when asked what she would look for prior to placing a student in a rapid advanced class, Sarah responded:

Definitely strength in math and language arts because language arts in the high schools are going to be so much a part of how they succeed in science and socials and with math, just the math in high school is very complex, so they need to have a really strong foundation to be successful.

When making decisions, one teacher covered the letter grades achieved in arts and music so as not to be influenced by what she referred to as "extraneous" or "non-academic" subjects. "I don't want to look at the music and the art. I'm looking at the other stuff." She further explained, "And so I think, just to make it easier for me, I would hide those two and just look at the top four." Violet concurs that when considering a record card, she too, ". . . would always check the language arts and math for an indication as to where to place them, and only later look at the others." When asked to explain why Violet reasoned:

Because language arts is so important for understanding other subjects and if they haven't got a grasp of maths, that can hold them back really a lot in secondary [...], So that's why I look at these two first because it indicates how they're going to do in other subjects.

During the task, high achieving grades in core subjects such as language arts were more likely to elicit strong reactions from teachers. This was particularly the case for the record card of a fictional student "Jamie Nelson" who achieved an A+ in language arts. While a middle range student, the A+ in that subject created considerable confusion among seven teachers trying to determine his suitable classroom placement. Some teachers were reluctant to place him in a regular class due to his high standing in language arts. For Rachel, Jamie's high level of achievement in that subject signified his ability to handle the workload in enriched classes, so she placed him in an advanced classroom. Peter, Stephanie, and Vanessa were more apprehensive, given his average achievement in other areas, and found themselves wishing for an alternative placement that would "challenge in one particular area rather than a complete advancement" but would not, as Vanessa said, put him at risk of "getting in over his head."

Just as an exceptionally high grade in a core subject prompted additional attention, so too could a lower grade in language arts or math. Lower grades achieved in these subjects were more likely to lead teachers to consider a remedial class placement, whereas lower grades in other subjects, such as science and history and, particularly, art and music, did not evoke as much concern. As Vanessa explained:

Students recommended for learning assistance are usually struggling across the board, but if their language arts and math are shaky, [...], I think that if you are going into grade

level with a C, your chances of passing the next grade level without something changing is very slim.

Some teachers' beliefs and values in relation to what they perceived as best socially for the student also played a role in their decision-making. For example, Sarah found it difficult to place students without knowing them, because she believes that "some kids need to be in a situation where they don't have to be super academic even if they could be because they're so low socially, they might not benefit from having a lot of academic pressure." For teachers like Sarah, a student may be academically competent, but, if the student is perceived as less socially adept than peers, a teacher may be less inclined to place them in high achieving classrooms. Other teachers expressed similar views, particularly with respect to ESL students, indicating how teachers' perceptions in relation to external student attributes such as race, gender and EAL status affect teachers' decision making.

External Student Attributes

A third overarching theme that emerged from the interviews was the role that teachers' perceptions of students' external attributes outside of their academic achievement (gender, ethnicity, EAL status) played in their decision-making. Various researchers (Peterson, Rubie-Davies, Osborne, & Sibley, 2016; Author 1 & 2, 2012) have suggested that one reason teachers hold different expectations of learners from different groups is due to implicit stereotypes, referring to a general belief about a group of people where certain traits or attributes are associated with that group (Greenwald & Banaji, 1995), and that these perceptions could either privilege or deny opportunities for certain groups of learners. For example, it was interesting that the four teachers who did place Jamie in an advanced class attributed their decision to the

priority they placed on language arts and the belief that "boys were less likely to achieve in language arts." For these reasons, Jamie's A+ in language arts was considered "exceptional" and warranted his placement in a rapid advanced class (Author 1, 2014b). The same comments were not made about Jamie's female counterpart with the corresponding record card, and three of the four aforementioned teachers placed her in a regular classroom.

Likewise, a student's EAL status could also play a role in teacher's placement decisions. While science was not regarded as a "core subject", the fact that Stanley had only earned a C in the course combined with his EAL student status elicited more concern from teachers like Peter and Nicole. Nicole notes:

Brooks has probably been speaking English all his life [...] whereas Stanley probably will need more assistance. It's already kind of a shock to go into grade seven to step into a new school and also be put in with the regular grade eights when you've been an ESL student since grade four. Supplementary might be a good way to give them a step up.

Both teachers rationalized their decision to place Stanley in a lower level class than his academically identical non-ESL, non-Aboriginal peer "Brooks Grayson", based on Stanley's status as a student for whom English is a second language.

Teachers in the study took more time and had less confidence in making decisions about the record cards of ESL (English Second Language) learners (see Author 1, 2014) or Indigenous learners (see Authors, 2012), groups that differed from their own. Sometimes teachers were impressed with the achievement of these students due to a belief that they had achieved despite their linguistic or Indigenous status. Other times teachers were hesitant to place a student at an appropriate level because they assumed language or family background might interfere. For

example, the marks of ESL students elicited skepticism due to the belief that grades may have been inflated specifically because of their status, as Janet's comment reveals:

I think the ESL kids struggle...and if these are true marks? Sometimes teachers want to encourage an ESL, and you can tell the kids smart, but it's a language issue. Sometimes teachers will put it up [student marks] just to encourage them.

Low-achieving ESL students were more likely to be placed in lower level classrooms, and more attention was paid to the grades achieved in language-based courses for ESL learners than their non-ESL peers (see Author, 2014b), as the following statement by Janet illustrates, "And I think it's pretty hard to be an A student when you are ESL in things like language arts, math not so much". Given this, ESL students may be at greater risk of remedial classroom placement despite their ability because teachers misattribute the lower achievement outcomes of ESL students to learning difficulties rather than language comprehension (Marx, 2006). Researchers (Author, 2015; Kubota & Lin, 2006; Rose, 2012) caution that apprehensive teachers may misplace EAL students in remedial classes, therefore, more comprehensive understanding of the complexities among English language learners is necessary, to ensure the decisions teachers make about linguistically diverse students are accurate (Tangen & Spooner-Lane, 2008).

Significantly, Veronique, the only teacher in the study to place all fictional students correctly, attributes both her ESL teacher training as well as being bilingual herself as contributing to her correct placement decisions. She suggests that it may be the assumptions some monolingual teachers hold about ESL students that make them inclined to undervalue the academic achievement of ESL learners.

She states,

I have the tendency to think that people that are multilingual can have an advantage on other students. They are able to play better with words and [...] they have a better meta-linguistic awareness. [...] But I think they [monolingual teachers] are probably just going on the stereotypes, assuming that because they're ESL, they're not able to achieve. [...] Maybe they just assume that they'll be ESL forever.

In a similar vein, some teachers perceived certain groups of students as expending greater effort than others, imputing meaning to the grades assigned to students as a reflection of the student's attributed ethnicity:

I think another part of it is that this might be, well, it's an Asian student of some sort. Korean or Japanese. There's a lot of emphasis on doing well and exceeding. Working really, really hard, lots of extra study, so I guess [...] culturally looking at it, I believed he would be able to handle the rapid placement.

While some teachers were positively influenced by the stereotypical perception of Asians as being "model minorities," they felt that a teacher's perception of work ethic could also be detrimental to the student if the student was perceived as spending too much time and too much effort to achieve certain grades. Olivia remarked:

Well, because the family, Korean parents, once again I'm generalizing, often put great pressure on their kids to achieve. They expect As and Bs. And when kids don't achieve they're often very hard on them. I've seen this time and time and time again.

Some teachers directly acknowledged how a student's background or ethnicity could influence their perceptions and potentially influence their placement decisions. When asked to comment

why she is so pleased with the high grades on the fictional record card of Aboriginal students, Sarah notes,

You don't really know why some kids are successful or not academically but, the general trend I think [is that] our society tells us [...] that Aboriginal kids aren't necessarily successful academically, so when you see it, you're like, "Oh great!" You know? You kind of notice it maybe more than an Asian kid who our society tells us is always academically successful right? So maybe it stands out more.

Such comments are indicative of the role teachers' perceptions of student attributes could have on teacher decision making if teachers' remain unaware of how internal biases may influence teacher judgment.

Discussion

While our study confirms what McMillain (2003) says about decision-making, it extends it further because McMillian's study does not take into account the consequences of bias. In this study, the fictional record cards of ESL and Indigenous learners were subjected to more scrutiny than record cards representing non-Indigenous, non-ESL learners (Author 1, 2014; Author 1 & 2, 2012). This was especially true in the case of high grades achieved by ESL students. Some teachers in our study felt that other teachers had the tendency to be "overly generous" when it came to evaluating ESL students. This study demonstrates how teachers' beliefs and values in relation to academic achievement, subject matter, and student behavior may also be intertwined with teachers' unchecked stereotypes and assumptions about specific groups of learners (Author 1 & 2, 2009; 2012; Author 2014; 2015).

Teachers play an important role in motivating students, building self-esteem and shaping their beliefs in terms of potential academic prospects (Gershonson, Holt, & Papageorge, 2016). While for teachers the grade may primarily be seen as an instrument of evaluation of student learning, for the student, the grade is not only a reflection of academic success but also represents "a criterion of his [sic] self assessment, serving as a kind of personal quality [...] an important personal characteristic (Sobkin & Kalashnikova, 2012, p.74). Yet studies indicate that teachers are more inclined to positively assess students who mirror similar values, belief systems, cultural norms, and group identity (Brophy & Good, 1974; Ford, Grantham, & Whiting, 2008; Gershonson, Holt, Papageorge, 2016, Karakowsky & Seigel, 1995; Bianco et al., 2011). This is a particular issue, given that while Canadian schools have become increasingly diverse, this diversity is seldom reflected among teachers who remain predominately from the dominant group (Ryan et al., 2009).

As Canadian schools become more diverse, teachers will need to understand how to adequately support learner diversity. Evaluating students based on the grades received for prior school performance figures prominently in a teachers' work. Despite this, our participants were apprehensive when asked to make decisions regarding student placement without actually meeting the student or having access to the opinions of their colleagues. For teachers in the study, access to a former teacher's anecdotal comments was useful for providing insight into the former teacher's assessment criteria (or lack of it). Fifteen of the twenty-one teachers in the study questioned the validity of grades provided by teachers with whom they had no prior experience.

It is illustrative of the confusion that grades engender for teachers that they criticized grading as of limited value because grades might reflect more than academic ability and because

grades might not reflect other attributes these teachers regard as important. Teachers who declared that grades alone were limiting because they did not include consideration of a student's work ethic or leadership potential were often those who declared that grades were limiting, precisely because of the possibility that grades might have been based upon factors other than a student's academic ability. This contradiction highlights the inconsistencies among teachers' interpretation of grades.

Some teachers were critical of basing student placement on grades alone, out of concern that grades were overly influenced by external factors such as student's status or perceived work effort. Other teachers were concerned precisely because they felt the grades represented on record cards did not take such factors into account. These teachers felt that, while grades might represent a student's academic aptitude, they did not reflect other factors they considered important, such as social skills or ability to co-operate with their peers. However, because determining what qualities constitute "good" social skills or social maturity is highly subjective, these teachers were also apprehensive when it came to making placement decisions without having insight into former teachers' beliefs and values. Some teachers were critical of basing student placement on grades alone, out of concern that grades were overly influenced by external factors such as student's EAL status or perceived work effort. Other teachers were concerned because they felt the grades represented on record cards did not take such factors into account. When asked to consider the grades on a student record card, one teacher remarked, "Is one teacher looking at effort and the other one only looking at a finished product?" It seems some teachers are influenced by how well they see their personal values reflected in the anecdotal comments made by the student's previous teacher.

This study confirms that teachers make disparate judgments in assigning grades and impute disparate meanings to the grades assigned by their peers. In addition, it illustrates how such disparity may be more pronounced for grades assigned to specific groups of learners. Given the strong reliance on high school grades for recommendation into higher education and scholarship education (Smith, 2003), the most important implication of this study is that considerable work is needed to prepare teachers to make judgments about students' work, and to recognize the injustice of imputing their own meaning to the grades assigned by others. Establishing clear performance standards and using exemplars of student work that are aligned with those exemplars can address these issues.

Recommendations

Veronique was the only teacher who placed all 24 student record cards accurately according to three different levels of achievement (supplementary learning assistance; regular grade eight, rapid advanced). She acknowledged that in addition to her bilingual status and ELL training, a key to her success was a program she attended designed to encourage teachers to critically reflect on the influence their values and biases may have upon their teaching. She notes that:

In that program what we did was mainly focus on doing introspection and going to the core of our beliefs and seeing how we all have judgments and stereotypes and we are never neutral. [...] I think what they [the instructors] did was focus on making us find out our values, our beliefs, why we have them, where they come from, and then challenge these and open them up. [...] I think that made a really big difference.

While some teachers may embrace the challenges that arise with increased self-awareness, others will remain resistant. Opportunities for teachers to engage in professional learning and professional practice devoted to assessment enable teachers to exercise judgment to avoid bias in assessment and to communicate effectively. Studies such as this are useful as they provide concrete evidence of how subconscious biases may influence the decisions teachers make about their students. If teachers are able to a) identify an issue and b) be convinced that they can do something about the issue, they will be more likely to modify the way they think about assessment.

Conclusion

Grades are used to summarize the set of assessments that teachers have made about students' performance in order to communicate the largest amount of information as simply as possible. Grades are also a substitute for the direct examination of the data that they summarize. Grades do not provide information about the methods or frequency with which the data were collected, the relative weighting, or any other element in the process used to derive the grade. As the comments of the teachers in this study indicate, grades can be influenced by and reflect dimensions that they were not intended to represent. When this occurs, they become unreliable indicators of performance, invalidating the inferences drawn from them. The comments of the teachers in this study highlight the ethical dimensions of the grading process. They found it difficult to make a recommendation regarding the students' placements on the basis of the grades because they were cognizant that the grades might reflect dimensions other than the students' performances. Although they did not use the term moral dilemma, they were quite conscious of what they believed was and was not acceptable for them to do.

Notwithstanding the many excellent attempts to address the challenges in the assessment, evaluation and reporting process in recent years, grades remain imperfect summary indicators and are especially problematic for making consequential decisions. Whether grades remain a staple of scholarly communication remains to be seen, but their elimination is unlikely to alleviate teachers' responsibility for making recommendations about students' opportunities based upon grades and other factors, or the moral obligations attendant upon such recommendations. With notable exceptions, teacher education and continuing professional education does little to help teachers recognize and address the ethical problems implicit in the decisions they make about students in relation to assessment. In other words, a missing dimension in teacher professional learning is the consideration of ethical issues. At the very least, focus on assessment, evaluation and reporting would benefit from attention on the moral components of the process and the responsibility inherent in making unbiased judgments.

References

Author 1 (2014a; b)

Author 1 & Author 2 (2008; 2012)

Alm, F., & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment, 20* (2), 132-150.

Auwarter, A. E., & Aruguete, M. S. (2008). Effects of student gender and socioeconomic status on teacher perceptions. *The Journal of Educational Research, 101* (4), 242-246.

Bennett, R. E. (2011). Formative assessment: *A critical review. Assessment in Education Principles Policy and Practice, 18* (1), 5-25.

Bennett, R. E., Gottesman, R.L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers: Judgments of students' academic skill. *Journal of Educational Psychology, 85*(2), 347-356.

Bianco, M., Harris, B., Garrison-Wade, D., & Leech, N. (2011). Gifted girls: Gender bias in gifted referrals. *Roeper Review, 33*(3), 170-181.

Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership, 49*(6), 15-18.

Bonesronning, H. (1999). The variation in teachers' grading practices: Causes and consequences. *Economics of Education Review, 18*, 89-105.

Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration, 47*(5), 609-629.

Boyatziz, R.E. (1998). *Transforming Qualitative Information*. Beverly Hills, CA: Sage.

Brackett, M. A., Floman, J. L., Aston-James, C., Cherkasskiy, L., & Salovey, P. (2013). The influence of teacher emotion on grading practices: A preliminary look at the evaluation of

- student writing. *Teachers and Teaching*, 19 (6), 634-646.
- Brookhart, S. M., Guskey, T.R., Bowers, A.J., McMillan, J., Smith, J.K., Smith, L.F., Stevens, M.T., & Welsh, M.E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86 (4), 803-848.
- Brookhart, S.M. (2004). *Grading*. Upper Saddle River, NJ: Pearson-Merrill-Prentice Hall.
- Brookhart, S. M. (1994). Teachers' grading: Practice and Theory. *Applied Measurement in Education*, 7(4), 279-301.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30 (2), 123-142.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10 (1), 35-36.
- Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt, Rinehart and Winston, Inc.
- Canadian Teachers' Federation. (2012). *Status of Women: Selected key issues affecting education*. Ottawa, ON: Canadian Teachers' Federation.
- Cheng, L. & Sun, Y. (2015). Teachers' grading decision-making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12 (2), 213-233.
- Clark, M., Thompson, P., & Vialle, W. (2008). Examining the gender gap in educational outcomes in public education: Involving pre-service school counselors and teachers in cross-cultural and interdisciplinary research. *International Journal of Advanced Counselling*, 30, 52-66.
- Colnerud, G. (1997). Ethical conflicts in teacher. *Teaching and Teacher Education*, 13, 627-635.
- Constas, M.A. (1992). Qualitative analysis as a public event: The documentation of category

- development procedures, *American Educational Research Journal*, 29, 2, 253-266.
- Cross, L.H., & Frary, R. B., (2009). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.
- DeBoer, B., Anderson, D., & Abdulaziz, E. (2007). Grading styles and instructor attitudes. *College Teaching*, 55(2), 57-64.
- Dunkake, I., Kiechle, T., Klein, M., & Rosar, U. (2012). Good looks, good grades? An empirical analysis of the influence of students' physical attractiveness on grading by teachers. *Zeitschrift für Soziologie*, 41 (2), 142–161.
- Fischbach, A., Budson, T. G., Preckel, F., Martin, R., & Brunner, M. (2013). Do teacher judgments of student intelligence predict life outcomes? *Learning and Individual Differences*, 27, 109-119.
- Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children*, 74, 289-206.
- Georgiou, S., Christou, C., Stavrinides, P., & Panourea, G. (2002). Teacher attributions of student failure and teacher behavior toward the failing student. *Psychology in the Schools*, 39(5), 583-595.
- Greenwald, A.G. & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102 (1), 4-27.
- Guskey, T. (2011). Five obstacles to grading reform. *Educational Leadership*, 69 (3), 17-21.
- Hinnant, J. B. O'Brien, M., & Ghazarin, S. R., (2009). The longitudinal relations of teacher expectations to achievement in the early years. *Journal of Educational Psychology*, 101, 662-670.

- Hoge, D.R., Hanson, S.L., & Smit, E.K. (1990). School experiences predicting changes in self-esteem of sixth- and seventh-grade students. *Journal of Educational Psychology*, 82 (1), 117-127.
- Imperial, P. (2011), *Grading and Reporting Purposes and Practices in Catholic Secondary Schools and Grades' Efficacy in Accurately Communicating Student Learning* (Doctoral dissertation, University of San Francisco, USA). Retrieved from <http://repository.usfca.edu/cgi/viewcontent.cgi?article=1002&context=diss>
- Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York: Oxford University Press Inc.
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131-155.
- Karakowsky, L., & Siegel, J. P. (1995). The effect of demographic diversity on causal attributions of work group success and failure. In M. J. Martinko (Ed.), *Attribution theory: An organizational perspective* (pp. 155-178). Delray Beach, FL: St. Lucie Press.
- Keller, T. (2016). Sticky assessments - the impact of teachers' grading standard on pupils' school performance. *Educational Studies*, 42 (5), 493-518.
- Kelly, S., & Carbonaro, W. (2012). Curriculum tracking and teacher expectations: Evidence from discrepant course taking models. *Social Psychology of Education*, 15, 271–294.
- Kohn, A. (2002). The dangerous myth of grade inflation. *Chronicle of Higher Education*, 49(11), B7.
- Kolb, K. J., & Jussim, L. (1994). Teacher expectations and underachieving gifted children. *Roeper Review*, 17(1), 26-30.

- Kubota, R., & Lin, A. (2006). Race and TESOL: Introduction to concepts and theories. *TESOL Quarterly*, 40, 471-493. doi: 10.2307/40464540
- Linn, R., & Miller, M. (2005). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson, Merrill, Prentice Hall.
- Marx, S. (2006). *Revealing the invisible: Confronting passive racism in teacher education*. New York: Routledge.
- Matteucci, M. (2007). Teachers facing school failure: The social valorization of effort in the school context. *Social Psychology of Education*, 10, 29-53.
- Matteucci, M., & Gosling, P. (2004). Italian and French teachers faced with pupil's academic failure: The norm of effort. *European Journal of Psychology of Education*, 19 (2), 147-166
- McCormick, T.M. (1994). *Creating the non-sexist classroom: A multicultural approach*. New York: Teachers' College Press.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd ed.) Thousand Oaks, CA: Sage Publications.
- McMillan, J (2003). Understanding and Improving Teachers' Classroom Assessment Decision Making: Implications for Theory and Practice, *Educational Measurement: Issues and Practice*, 22 (4), 34-43.
- McMillan, J.H. & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making*. Paper presented at the 2000 Annual Meeting of the National Council of Measurement in Education, New Orleans, LA.
- McMillan, J.H., Myran, S., & Workman, D. (2002). Elementary Teachers' Classroom Assessment and Grading Practices. *The Journal of Educational Research*, 95 (4), 203-213.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33, 49-64.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp.13-103). New York, NY: American Council of Education and Macmillan.

Mizala, A., Martinez, F., & Martinez, S. (2015). Pre-service elementary school teachers' expectations about student performance: how their beliefs are affected by their mathematics anxiety and student's gender. *Teaching and Teacher Education*, 50, 70-78.

Ontario Ministry of Education. (2010) *Growing Success: Assessment, Evaluation and Reporting in Ontario Schools* (1st ed., covering Grades 1 to 12). Retrieved from <http://www.edu.gov.on.ca/eng/policyfunding/growsuccess.pdf>

Peterson, E.R., Rubie-Davies, C., Osborne, D. & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, 42, 123-140.

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.

Protz, T. (2013). Variation in grading practices-subjects matter. *Education Inquiry*, 4 (3), 555-575.

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26, 1372-1380.

Ready, D.D., & Wright, D.L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. The role of child background and classroom context. *American Educational Research Journal*, 48, 335-360.

- Reyna, C. (2000). Lazy, dumb, or industrious: When stereotypes convey attribution information in the classroom. *Educational Psychology Review*, 12, 85-110.
- Rose, M. (2012). The silent apartheid as the practitioner's blind spot. In K. Price (Ed.), *Aboriginal and Torres Strait Islander Education: An Introduction for the Teaching Profession* (pp. 66-81). Port Melbourne, Vic: Cambridge University Press.
- Rubie-Davies, C. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology*, 80, 121-135.
- Rubin, I.J. & Rubin, I.S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). London: Sage.
- Ryan, J. Pollack, K. & Antonelli, F. (2009). Teacher diversity in Canada: Leaky pipelines, bottlenecks, and glass ceilings. *Canadian Journal of Education*, 32, 3, 591-617.
- Sarra, C. (2011). *Strong and smart-towards pedagogy for emancipation: Education for first peoples*. New York: Routledge.
- Scott, S., Webber, C.F., Aitken, N., & J. Lupart (2011). Developing teachers' knowledge, beliefs, and expertise: Findings from the Alberta student assessment study. *The Educational Forum*, 75(2), 96-113.
- Shepherd, M. A. (2011). Effects of ethnicity and gender on teachers' evaluation of students' spoken responses. *Urban Education*, 46 (5) 1011-1028.
- Siegle, D., & Reis, S. (1998). Gender differences in teacher and student perceptions of gifted students' ability and effort. *Gifted Child Quarterly*, 42, 39-47.
- Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational measurement: Issues and practice*, 22(4), 26-33.
- Stanley, G. & Baines, L. (2001). No more shopping for grades at B-mart: Re-establishing grades as indicators of academic performance. *The Clearing House*, 74(4), 227-230.

- Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69 (5), 797-811.
- Sudkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgments of student's academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, 743-762.
- Sun, Y., & Cheng, L. (2013). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326-343.
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings': An attempt to verbalise and discuss teachers' internalized grading criteria. *European Physical Education Review*, 20 (2), 199-214.
- Tangen, D. & Spooner-Lane, R. (2008). Avoiding the deficit model of teaching: Students who have EAL/EAL and learning difficulties. *Australian Journal of Learning Difficulties*, 13 (2), 63-71.
- Thorndike, R. (2005). *Measurement and evaluation in psychology and education (7th ed.)* Upper Saddle River, New Jersey: Pearson Education, Inc.
- Timmermans, A.C., de Boer, H., & van der Werf, M.P.C. (2015). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Sociology Psychology Education*, 19, 217-240.
- Tollefson, N. & Chen, J. (1988). Consequences of teachers' attributions for student failure. *Teaching and Teaching Education*, 4 (3), 259-265.
- Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R.W. (2010). The implicit prejudiced attitudes of teachers: Relations to teach expectations and the ethnic achievement gap. *American Educational Research Journal*, 47, 497-527.

- Weiner, B. (1974). *Attribution motivation and attribution theory*. Morristown, NJ: General Learning Press.
- Weiner, B. (1985). An attributional theory of achievement motivation. *Psychology Review*, 92, 548-573.
- Wilson, R. J. (1996). *Assessing students in classrooms and schools*. Toronto: Allyn & Bacon.
- Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors* (Doctoral thesis, Virginia Commonwealth University, Richmond, Virginia). Retrieved from <http://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=3585&context=etd>
- Ziomek, R. L., & Svec, J. C. (1997). High school grades and achievement: Evidence of grade inflation. *NASSP Bulletin*, 81(587), 105-113.
- Zoeckler, L. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education*, 35 (2), 83-1-2.