

Validity of Grading Practices in Elementary Education

Shannon Thoendel

EDL 9950 Symposium on School Leadership

April 2, 2015

### **Abstract**

No Child Left Behind (NCLB) has required states to adopt academic standards and develop tests to measure student achievement. District, school, teacher, and student performance is ranked based on adequate progress towards academic standards achievement. The reliable measure to communicate and monitor student progress to stakeholders is teacher assigned report card grades. Report card grades are used for multiple purposes in education and can impact the educational welfare for students. It is well documented that teacher grading practices are inconsistent. Grading inconsistency is accentuated by a focus on grades that may lead to a negative impact on student learning, individual student characteristics, and the individual practices or beliefs of educators. The research demonstrates by focusing teacher grade reporting and assessments on the academic standards, validity in grading practices continue to increase and provide a clearer picture of student learning that is necessary to increase student achievement towards proficiency of the academic standards. Furthermore, the research demonstrates grading practices can be improved through teacher training and a well communicated school district grading policy.

## **Research Topic**

The pressure for successful performance on high-stakes tests has increased the necessity for valid measures of academic achievement that is accurately communicated to parents, teachers, and administrators. The most relied upon source of student performance communication is the assignment of report card grades by teachers. This generates the question: Do teacher assigned grades validly measure and communicate student academic performance?

## **Introduction**

### **History of Grading**

The history of grading suggests that before 1850, most schools grouped students of all ages and backgrounds into one cluster in a one-room schoolhouse. Grades were reported to parents orally during a home visit. Grading and reporting were unheard of in U. S. schools at this time (Guskey, 2013). In the late eighteen hundreds schools began to use formal progress evaluations. These were primarily narrative reports where teachers described the skills the student had mastered and where additional work was needed. The main objective of these reports was to communicate mastery of current level and readiness to move to the next (Edwards & Richey, 1947).

In the nineteenth and twentieth centuries, enrollment increased due to compulsory attendance laws. The number of US schools increased dramatically. Subject specific content increased along with the number of high schools (Gutek, 1986). Elementary schools continued to use narrative reporting for student performance, where high school teachers began to use percentages and other similar markings to communicate achievement (Kirschenbaum, Simon, & Napier, 1971).

The shift to percentages was gradual and seemed natural due to the increased demands on high school education. In 1913, Daniel Starch and Edward Charles Elliott challenged the reliability and accuracy of percentages. The research found wide differences in assigned percentages to identical English papers. The teachers focused on different elements of the paper. Thirty different percentage grades were assigned to a single paper and scores had a range of more than 40 points (Guskey, 2013). The study was repeated with geometry papers and the researchers found even greater variation in grading. The math papers ranged in grades from 28% to 95% (Guskey, 2013). History has shown that the validity in teacher grading practices has been questioned for over one hundred years. This paper will examine the research that addresses grading validity concerns and outline what research indicates to have a positive impact on student assigned grades.

### **Purpose of Grading**

Grades are significant in American education systems. They are used to determine class placement, retention practices, college admissions, and scholarships. Today's report cards are used to sustain state funding, generate positive feelings between school and community, assist teachers in increasing students' self-esteem, used as a reward for student likeability, and create a chance to receive college funding (Stanley & Baines, 2004). Grades should simply reflect academic performance towards learning goals (Randall & Engelhard, 2010). The problem arises when grades are not just limited to communicating student achievement; they include self-esteem boosters, attitude, participation, and rewards (Stanley & Baines 2004).

The pressure for successful performance on what would be considered high-stakes testing has increased the necessity for valid measures of academic achievement that is accurately communicated to parents, teachers, and administrators. There is agreement in the education

community that teacher assigned grades can be a reliable and valid tool for communicating a student's academic progress. Guskey (2007), found that diverse stakeholders perceive validity of achievement indicators differently. Administrators view state, district, and national standardized assessments more valid, where as teachers perceived classroom observations and homework more valid than administrators.

Guskey (2007), also investigated the purpose for grading by teachers, parents, and students. His findings can be classified into six broad categories: (1) to communicate the achievement status of students to parents or others, (2) to provide information for student self-evaluation, (3) to identify certain pathways or instruction in education, (4) to provide learning motivation and incentives for students, (5) to evaluate the effectiveness of instructional programs, and (6) to provide evidence of student effort or inappropriate accountability. In 2007, Guskey also sought to determine similarities and differences in all three stakeholders: parents, students, and teachers. Guskey wanted to determine their perceptions on the purpose of grading and reporting. Stakeholders ranked the six major purposes of grading mentioned above. Each group ranked "communication to parents" and "feedback to students" as the most important purpose for grading. The two least important purposes were "evaluation of instructional programs" and "lack of effort and accountability".

### **Standards-Based Grading**

Traditionally, grades have been constructed on assessment methods designed by educators and are based on comparing an individual student with a group. For grades to have meaning, there must be a clear understanding and a point of reference to compare student

achievement. With the adoption of state and national student performance standards, grades that are based on standards will allow educators to use a criterion-referenced approach in assessing student achievement (O'Connor, 2009).

To combat grading issues and misinterpretations of grades, Ken O'Connor (2009) outlines a criterion-referenced grading system based on standards. Standards-based grading is where grades are strictly based on learning outcomes or performance standards that create a clear portrait on what students should know and be able to do. These standards create greater equity on learning outcomes for all students through consistent communication about student achievement among stakeholders (Schmoker, 2000).

A strong standards-based grading system includes many attributes. It views grading as a process. Quality criterion-referenced performance standards should be the reference point to determine student grades. Value or judgment attributes should be limited, and not all student work samples should be included in grades. Students should be allowed to work towards mastery, keeping grades written in pencil for the possibility of improvement. All grading procedures should align with learning goals. Standards-based grading is a teaching and learning process that involves properly recorded evidence of student achievement (O'Connor, 2009).

### **No Child Left Behind**

No Child Left Behind (NCLB) was assigned into law by President Bush on January 8, 2002.

With the implementation of this law came more stringent local, state, and federal accountability measures for school systems. Under NCLB, states were required to develop content and academic achievement standards. To measure how well all students in the state are acquiring the skills defined by the content standards, state assessment tests were also a requirement.

Annually, students in grades 3 through 8 are tested in the areas of reading, mathematics, and science on the academic achievement standards. Student performance on the state mandated tests is ranked into two high levels and a third lower level. All students are expected to reach a level of proficiency. NCLB requires the total student population and specified subgroups to meet “adequate yearly progress” (AYP) by reaching the proficient level, which is defined by attaining the two higher level rankings on the mandated state tests. If schools and districts fail to meet AYP for two or more years, they are then classified as schools “in need of improvement” and face such consequences as school transfer options, supplemental services, replacement of staff or administration, or a plan of restructuring (Great Schools, 2015).

NCLB directives hold states and schools more accountable for student academic results. State mandated tests have high-stakes consequences which cause these tests to grow in significance. Because the tests are standardized assessments given under uniform conditions, they are considered to be a meaningful basis for evaluating performance (Coladarci, 1986).

## **Summary of Findings**

### **The Effects of Grading On Students**

Grades have been linked to having a strong and lasting impact on a student’s attitude, behavior, and motivation for learning (Brookhart, 1994). Researchers refer to three consistent effects that arise when an emphasis is placed on the importance of letter or number grades.

Assigning an arbitrary letter or number grade tends to: (1) reduce the student’s interest in actual

learning, (2) increase the frequency of students choosing the easiest task, and (3) lessen the quality of the student's thinking (Kohn, 1999).

Research has shown that “focus on grades” and “focus on learning” are opposite of each other (Kohn, 1999). Kohn, moreover, states that when students are told they need to know something, they tend to lose interest. If a student is focusing on the grade and there is pressure to receive the highest grade, the easiest intellectual path will produce this result. The end result being a reduction in thinking due to lack of interest and poor intellectual exploration.

Teachers do understand the importance of grades for students and parents. Students who receive good marks may get paid by parents and get to attend reward celebrations. Good grades build student self-worth. These are all factors that impact a teacher's ability to assess based on only achievement (Randall & Engelhard, 2010).

### **Common Grading Design**

More detailed reporting methods such as checklists or narratives have their disadvantages too. These reporting methods offer more specific information on student achievement, but take greater amounts of time to prepare. More detailed reporting systems that increase the analytical process, are more likely to lead to subjectivity that influences grades (Ornstein, 1994). However, not all subjectivity in grading is bad. The teacher knows the students the best, understands the limitations of the work completed, and has knowledge of the progress made in class. This may produce a more accurate picture of the students' current academic performance (Brookhart, 1993).

Kohn (1999) would argue that several research studies found that students given numerical grades were far less creative than those students who received qualitative feedback. A combination of comments and numerical grading did not help attain high achievement. Students



achieved the highest when comments were given instead of letter or number grades (Kohn, 1999).

According to Reeves (2011), traditional grading systems are inconsistent inaccurately conveying how students perform in relations to learning standards. Grades are influenced by a variety of factors that include effort and behavior. In traditional systems that use numbers or letters to represent achievement, grades only slightly relate to performance on high-stakes external assessments. Today's education system demands large amounts of accountability. This is evident in NCLB. Grading systems must reflect an accurate measure of students' learning towards the content standards. Many school districts have initiated standards-based grading practices in hopes of gaining a more reliable measure of student mastery of adopted learning standards.

One goal of standards-based grading systems is to remove subjectivity in grading, thus providing valid information about student learning (Hardegree, 2012). This was verified in a 2012 study by Hardegree. The study revealed that standards-based report card grades accurately correlated to the students' performance on a required high-stakes standards-based state test. In fact, the teacher assigned standards-based report card grades, exceeded the requirements of proficiency on the state test. By focusing teacher grade reporting and assessments on the standards, educators continue to increase validity in grading practices and provide a clearer picture of student learning that is necessary to increase student achievement (Hardegree, 2012).

### **Grading Validity Factors**

Reeves (2008, 2011) expressed it is not new curriculum, replacement of a principal or teachers, or great technology that will improve schools; it is simply the need for a better grading system. He further stated that policies have to be set that require teachers to calculate grades

solely on academic performance. Even teachers who agree with grading systems that only reflect academic performance struggle to keep non-achievement factors out of their grading practices. In reinforcement of this statement, eighty-one percent of teachers polled and seventy percent of students agreed or tended to agree with the statement that achievement should be reported separately from other factors (Cross & Frary, 1999). Even though a high majority of the teachers expressed a belief in the statement, their actions do not indicate agreement. Seventy-two percent of the teachers in the same study indicated they raised the grade of low-ability students based on other factors than achievement (Cross & Frary, 1999). Similar results were found when teachers professed to adhere to grading practices that were aligned with best practice research on grading; however, when they replied to a grading survey their responses indicated differently (Steidinger, 2011).

A student's characteristics have also been found to impact the validity of grades. A teacher's perception of a student's behavior can significantly influence judgments of his or her academic performance. Four major factors are considered by teachers when assigning a final grade: Student academic achievement, ability, behavior, and effort (Randall, & Engelhard, 2010; Südkamp, Kaiser, & Möller, 2012). A teacher's perception of a student's behavior can significantly influence the reporting of a student's academic performance (Hills, 1991). Even the neatness of a student's handwriting can influence a student's grade (Sweedler-Brown, 1985; Steidinger, 2011).

Brookhart (1993) demonstrates how value judgement and subjectivity can impact a student's grade. Teachers in a study were directed to assign a grade in two different situations. An average Algebra I student recorded grades on two tests for the grading period. On the first test he achieved an F and on the second test he achieved a low D. The teachers were asked to

assign the student an overall grade for the period. The choices were an overall grade of an F based on the average of the two tests or an overall D because improvement of performance was demonstrated. Seventy-three percent of the teachers chose the D. The second situation was similar, except this Algebra I student achieved a B on his first test and a low A on the second test. The choices for a final grade were an overall grade of B which was the average of the two test grades or an A with the consideration there was improvement. With an identical percentage, this time the teachers chose B as the grade (Brookhart, 1993).

In 2010, Randall and Engelhard shared their examination of factors that impact borderline decisions in grade assignment. A student with a report grade of sixty-nine percent that demonstrates low achievement and low ability, but offers good behavior with high effort, on average receives a grade of seventy-seven percent. This solidifies the thought that teachers reward lower achieving students at a higher grading rate due to good behavior and effort. This was not only true for low achieving students. Students that were on the borderline of achieving an A or a B, were consistently receiving the higher grade when their effort and behavior were excellent. Regardless of ability, students' grades improve with good behavior or effort (Randall & Engelhard, 2010).

### **Correlation of Teacher Performance Judgements and High-Stakes Testing**

When teachers are asked to predict student performance on a standardized test, their judgement is not always found to be accurate. Teacher judgement accuracy has been found to differ between low and high achieving students. Teachers were able to accurately judge student responses for three quarters of the time on individual test items, but accuracy declined within subject subtests. Teachers were the least accurate when judging low achieving students, increasing accuracy for the high achievers. This was determined to be true because the high

achieving student would answer a large amount of the questions correctly and the teacher assumed automatically the high achiever was efficient. This thinking would not apply to the lower achieving student since the student would answer many of the questions wrong. The implications of this are the students in greatest need of accurate performance appraisal and education assistance are the students who are at the greatest risk of being misjudged and in return, may lack the support they need to increase testing performance (Coladarci, 1986).

### **Practices to Improve Grading**

Regardless of teacher training on assessment, when assigning grades, most consider ability and effort (Brookhart, 1993). Forty-seven percent of elementary teachers reported using ability when assigning grades, especially when making borderline grading decisions (McMillan, Myran, & Workman, 2002). When a classroom teacher decides what factors will impact grading, one's personal training, official grading policies, and perceived consequence of assigned grades influence the outcome (Randall & Engelhard, 2010).

An exercise that has produced higher levels of consistency in grading practices is the training of personnel, directly involved in the grading process, on the school district's philosophy and procedures for assigning grades (Mehring, Parks, Walker, & Banikowski, 1991). This conclusion was decided through a study that sought to define whether training on grading procedures would impact the degree of variability for grade assignment. Elementary principals and teachers in a large suburban school district were asked to complete a questionnaire before and after they received training on their current report cards which contained two marking systems.

In the case of both marking systems, pre-training results indicated that the majority of the participants felt that the marking systems were unclear, and they were dissatisfied with the

systems. Post-training, a difference in perception was collective. The majority of survey participants indicated the marking systems clearly communicated the students' present levels, and they were in favor of retaining the current grading systems. Further noted, before training, vast differences in interpretation of both marking systems were evident. Greater consistency in assigning grades was achieved when teachers implemented a shared philosophy and defined criteria for assessing student achievement (Mehring et al., 1991).

### **Implications of Findings and Application Metropolitan Omaha Education Consortium**

Implications for the Metropolitan Omaha Education Consortium, MOEC, can be drawn from the results of grading research. Two common themes appeared in the research (1) teacher beliefs and practices contribute to validity issues for grading and (2) teacher training can increase validity in grading. In light of the research, combined with the mission of MOEC, to bring support in a collegial environment for all surrounding metropolitan school districts in the areas of teaching, research, and service to improve education, should not MOEC members work together to achieve equal opportunity in grading for all students?

The research shows that teacher assigned grades serve many purposes. This includes communication to stakeholders, class placement, retention practices, college admissions, and scholarships measurements. Student grades are also used to generate positive feelings between the school and community, perceived to assist in increasing a student's self-esteem, and used as a reward for students (Stanley & Baines, 2004). Grades are a significant part of a student's education career, but yet the research solidifies there is little consistency or validity in their assignment.

According to the Nebraska Department of Education 2013-2014 State of the Schools Report (2015), local MOEC school districts reached a mobility rate as high as ten percent. If ten

percent of the student population is in transition, a good educational practice would be to provide consistent communication of student achievement. Research has revealed that student transfers from one education institution into another can cause academic regression (Straits, 1987). Student academic regression could be identified and addressed immediately, if performance scales and reporting system were consistent between educational institutions.

Inconsistent and unreliable grading practices can be immediately addressed through MOEC collaboration. MOEC experts should work together to create professional development and teacher preparatory classes that address (1) teacher beliefs and practices that contribute to validity of grading and (2) teacher training to increase validity. Moreover, MOEC can facilitate common research based grading practices across the metropolitan school districts, which could then extend further through the state of Nebraska increasing the chances of equal educational opportunities for all students.

### **Staff Development and Teacher Training**

Variability among teachers' judgements of student performance can be significantly influenced by the teachers' assessment beliefs and practices. Teacher extensive training in the area of grading and assessment contributes to the reliability of grading practices (Sweedler-Brown, 1985). The higher learning institution and MOEC partner, University of Nebraska at Omaha, has a liability to properly train emerging educators. Placing pre-service teachers into the education profession without proper grading knowledge, would be a disservice to the students and the local school districts that depend on hiring quality professionals.

MOEC states that staff development and assistance to beginning teachers is a high priority issue. Before assistance and staff development can be implemented, the MOEC Assessment, Staff Development, and Curriculum Task Forces need to formulate a joint vision

and philosophy for student grading. According to the research, a standards-based grading system strongly correlates with student achievement on high-stakes tests. No Child Left Behind states that one hundred percent of students must reach a proficient level on a state selected standardized test that measures a student's proficiency on state adopted standards. School curriculum adoption materials align with the state adopted standards, teachers are held accountable to teach to the standards, and student achievement is ranked by school, state wide, based on achievement of the standards (Hardegree, 2012).

Currently many school districts in the MOEC community report student achievement on assignment of a letter or number grade based on a percentage of a number correct on assigned homework and tests. The purpose of grading should be to communicate information about how well students have learned content (Guskey, 2013). A single letter or number does not communicate an area of deficiency or strength to stakeholders. The research communicated there is great variation in teacher assigned grades. Grading bias and variance may be minimized by implementation of a integer grading system or standards-based grading system that aligns with the Nebraska State Reporting System of below, proficient, or advance achievement on the Nebraska State Accountability Test (Guskey, 2013; Steidinger, 2011; Hardegree, 2012).

A joint MOEC effort to improve grading practices and increase teacher efficacy in grading would lead to a clearer vision and purpose for higher learning institutions and school districts trying to adopt research based grading practices. Some MOEC school districts have been met with great resistance from constituents when transitioning to a standards-based grading system. MOEC task forces can help communicate research, train, and support district efforts. This may come in the form of communication through research, op-ed articles, and extensive

coursework for CADRE Project Teachers, Teacher Academy Project, and Mentor Project in the area of grading.

### **Universal Grading Practices**

With mobility rates reaching as high as ten percent in the metropolitan area, consistency among school districts' grading practices is a concern (Nebraska Department of Education, 2015). There is no exact mobility data that quantifies student movement between state systems or out of state systems. Good educational practice should be consistent among all educational institutions. MOEC is a collaborative environment that could work together to create consistency of grading practices and a universal grading system. When one school district uses letter grades or percentages on a varying scale, or another district uses standards-based grading, it is difficult for educators, administration, and parents to correlate student performance. As a student transfers through a school system or across a school system, this inconsistency can negatively impact performance (Straits, 1987). Universal reporting would allow early diagnosis of regression and implementation of practices to alleviate dramatic decreases in student performance.

The MOEC community can use its combined expertise and influence to initiate state wide education reform on universal grading practices. This would be suggested in the form of a standards-based grading system that aligns with the State content and academic achievement standards. Schmoker (2000) explains, standards-based grading assesses student performance strictly on outcomes or standards; deriving a clearer picture of student knowledge and creating consistent communication among stakeholders.

Many states have proposed and implemented grading policies that are usually characteristics of standards-based grading. Such practices as a "no zero" grading policy or



policy that bands school districts from assigning a percentage grade lower than fifty percent support the concept that grading is a process leading towards student mastery. Implementation of these policies eliminates severe grading practices that perpetuate no chance of grade recovery for students (Guskey, 2004).

When grading practices are used in attainment of scholarships and job placement, inequivalent grading scales leave some students at a disadvantage. Grades are a universal symbol used to communicate evidence of student achievement. Because grades have implications for a student's future, MOEC and partnering school systems should engage in a robust effort to establish consistent and effective grading systems.

### References

- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *JEDM Journal of*
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology, 78*(2), 141-146.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53-72.
- Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues & Practice, 26*(1), 19-27.
- Hardegree, A. C. (2012). *Standards-based assessment and high stakes testing: Accuracy of standards based grading* (Doctoral dissertation, School of Education Liberty University).
- Hills, J. R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan, 72*(7), 540-45.
- Kohn, A. (1999). Grading is degrading. *Education Digest, 65*(1), 59.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*(4), 203.
- Mehring, T., Parks, C., Walker, K., & Banikowski, A. (1991). Report cards: What do they mean during the elementary school years. *Reading Improvement, 28*(3), 162-68.
- Ornstein, A. C. (1994). Grading practices and policies: An overview and some
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380.

- Reeves, D. B. (2008). *Effective grading*. Association for Supervision & Curriculum Development.
- Reeves, D. B. (2011). *Elements of grading: A guide to effective practice*. Bloomington, IN: Solution Tree Press.
- Steidinger, R. (2011). *A study to determine the concurrent validity between course grades and standardized assessment results* (Doctoral dissertation, Department of Educational Leadership East Carolina University).
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762.
- Sweedler-Brown, C. O. (1985). The influence of training and experience on holistic essay evaluations. *The English Journal, 74*(5), 49-55.

#### **Other Resources**

- Edwards, N. & Richey, H. G. (1947). *The school in the American social order; the dynamics of American education*. Boston: Houghton Mifflin Co.
- Great Schools. (2015). *No child left behind requirements for schools*. Retrieved from <http://www.greatschools.org/definitions/nclb/nclb.html>
- Guskey T.R. (2004). *Stability and change in high school grades*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. ERIC document retrieved from <http://files.eric.ed.gov/fulltext/ED490411.pdf>
- Guskey T.R. (2013). The case against percentage grades. *Educational Leadership, 71*(1), 68-72.
- Gutec, G. L. (1986). *Education in the United States: An historical perspective*. Englewood Cliffs, N.J.: Prentice-Hall.

- Kirschenbaum, H., Simon, S. B., & Napier, R. (1971). *Wad-ja-get?: The grading game in American education*. New York: Hart Pub. Co.
- Nebraska Department of Education. (2015). *State of the schools report 2013-2014*. Retrieved from <http://reportcard.education.ne.gov/>
- O'Connor, K. (2009). *How to Grade for Learning*. California: Corwin.
- Schmoker, M. (2000). Standards versus sentimentality: Reckoning-successfully-with the most promising movement in modern education. *NASSP Bulletin*, 84(620), 49-60.
- Stanley, G., & Baines, L. (2004). No more shopping for grades at B-mart. *Clearing House*, 77(3), 101-104.
- Straits, B.C. (1987). Residence, migrations, and school progress. *Sociology of Education*, 60, 34-43.

### **Author's Information**

Shannon Thoendel is a doctoral student at the University of Nebraska at Omaha. She has 23 years' experience in the education field as a classroom teacher, mentor, and administrator. The past eleven years she has been an elementary principal, with experiences in a high impoverished urban district to a rural district quickly transforming into a suburban system. Shannon currently serves as a kindergarten through sixth grade principal at Pine Creek Elementary School. Pine Creek Elementary is located northwest of the Omaha metropolitan area in the Bennington Public Schools District. Bennington Public Schools is one of the fastest growing school District in the State of Nebraska. Shannon has had the opportunity to provide guidance and leadership as the District has transformed from one elementary to three elementary facilities in nine years.