

Exploring the factors teachers consider in determining students' grades

Thomas R. Guskey & Laura J. Link

To cite this article: Thomas R. Guskey & Laura J. Link (2018): Exploring the factors teachers consider in determining students' grades, *Assessment in Education: Principles, Policy & Practice*, DOI: [10.1080/0969594X.2018.1555515](https://doi.org/10.1080/0969594X.2018.1555515)

To link to this article: <https://doi.org/10.1080/0969594X.2018.1555515>



Published online: 11 Dec 2018.



Submit your article to this journal [↗](#)



Article views: 75



View Crossmark data [↗](#)



Exploring the factors teachers consider in determining students' grades

Thomas R. Guskey^a and Laura J. Link^b

^aDepartment of Educational, School, and Counseling Psychology, College of Education, University of Kentucky, Lexington, KY, USA; ^bCollege of Professional Studies, Purdue University Fort Wayne, Fort Wayne, IN, USA

ABSTRACT

The purpose of this study was to investigate the specific factors teachers consider when assigning students' report card grades. Data were gathered from 943 K-12 teachers from five school districts in a southeastern state in the United States who completed the *Teachers' Grading Practices Survey*. Analyses focused on how teachers weigh different factors in determining report card grades, and if these factors and weights differ among teachers who teach at different grade levels and have different amounts of classroom experience. Results revealed statistically significant differences among teachers at different grade levels but no differences associated with teachers' years of experience and no interaction effect. Differences by grade level were evident in teachers' consideration of both cognitive and non-cognitive factors of students' performance. Implications are discussed for improving grading policies and practices, teacher education and teacher professional development.

ARTICLE HISTORY



Received 20 February 2018
Accepted 26 October 2018

KEYWORDS

Grading; assessment; teacher grade level; teacher years of experience; non-cognitive factors

The grades teachers use to describe students' performance in school and record on report cards have long been identified by the measurement community as prime examples of unreliable measurement (Brookhart, 1993; Stiggins, Frisbie, & Griswold, 1989). Numerous studies show teachers vary widely in the criteria they use in assigning grades (Adrian, 2012; Bailey, 2012; Cross & Frary, 1999; Duncan & Noonan, 2007; Grimes, 2010; Guskey, 2002, 2009b; Imperial, 2011; Liu, 2008; Llosa, 2008; McMillan & Lawson, 2001; McMillan, Myran, & Workman, 2002; Randall & Engelhard, 2009, 2010). Even in schools where established policies offer guidelines for grading, significant variation in grading practices remains among teachers who teach at the same grade level or in the same academic department (Brookhart, 1994, McMillan, 2001).

Despite their documented unreliability and highly varied use among teachers, grades remain the primary indicator of how well students perform in school. As such, they serve as the basis for making numerous important decisions about students (Brookhart et al., 2016; Guskey, 2015). Report card grades largely determine whether or not

CONTACT Thomas R. Guskey  guskey@uky.edu  Department of Educational, School, and Counseling Psychology, College of Education, University of Kentucky, Lexington, KY 40506, USA

An earlier version of this paper was presented at the annual meeting of the American Educational Research Association, San Antonio, TX, April 2017.

© 2018 Informa UK Limited, trading as Taylor & Francis Group

students are promoted from one grade level to the next. They also determine honour roll status, enrollment in advanced or remedial classes, special education services and college or university admissions (Brookhart, 1994; Brookhart & Nitko, 2008; Imperial, 2011). Because of the relationship between report card grades and educators' decision-making, students' academic opportunities may be affected when significant grading variation exists among teachers (Guskey, 2015; Link, 2018).

Recent empirical studies based primarily on surveys of teachers' grading practices reveal that most teachers use a combination of cognitive and non-cognitive evidence, primarily perceptions of student effort and teachers' professional judgment of students' ability, in determining grades (Brookhart, Guskey, Bowers, McMillan, Smith, J., Smith, L., Stevens, & Welsh, 2016; McMillan & Nash, 2000; Randall & Engelhard, 2010; Russell & Austin, 2010; Sun & Cheng, 2014; Svennberg, Meckbach, & Redelius, 2014; Yesbeck, 2011). This collection of evidence is then tallied by teachers in various ways into a single amalgamated, 'hodge-podge' grade (Brookhart, 1991, p. 36) that mixes achievement and other factors related to effort, behaviour, attitudes and improvement. The effects of this grading conflation are compounded in the context of current high-stakes assessment and accountability processes that are typically designed to measure only student academic mastery (Brookhart et al., 2016; Kolio-Keaikitse, 2012).

The purpose of this study was to investigate, with more specificity than earlier studies, the factors teachers consider when determining the grades they record on students' report cards. In this research, the term 'factors' is used to describe the sources of evidence teachers consider and the specific judgments teachers make in determining the grades assigned to students. Other researchers have referred to these factors as 'student evaluation techniques' (Gullickson, 1985) and 'grading dimensions' (Stiggins et al., 1989). Specifically, we sought to determine if these factors vary among teachers who teach at different grade levels and among teachers with different years of classroom experience.

As students mature and develop more sophisticated communication skills; especially in reading, writing, and speaking; teachers employ a wider variety of assessment techniques to gather information on students' achievement of learning goals (Brookhart, 1993; Gullickson, 1985; McMillan, 2001). It was hypothesised that the availability of more varied assessment formats would result in teachers at higher grade levels using different and more diverse types of evidence to determine students' grades. In addition, in recent years there has been increased emphasis in undergraduate teacher education programmes on assessment literacy and the use of more effective grading and reporting practices (Gareis & Grant, 2015; Popham, 2009; Yamtim & Wongwanich, 2014). It was further hypothesised that due to this emphasis, the grading policies and practices of newer teachers would differ from those of their more experienced colleagues (Deluca, Coombs, LaPointe, & Chalas, 2017; Deluca & Klinger, 2010; Link, 2018, Volante & Fazio, 2007).

Theoretical framework

Extensive research shows that teachers use a variety of sources of evidence related to both cognitive and non-cognitive factors in determining students' grades (Brookhart et al., 2016). In an early study of grading practices, Gullickson (1985) compared

teachers' use of 11 different sources of evidence, what he labelled 'evaluation techniques,' in assigning grades. These included four types of assessments (standardised objective tests, teacher-made objective tests, essay tests and oral quizzes), five student activities (class discussions, oral reports, student papers or notebooks, group or individual projects and laboratories) and two student behaviour categories (citizenship and behaviour displayed in school, and citizenship observed in the community). No explanation was offered, however, as to why these particular grading factors were chosen.

Gullickson (1985) found that teachers' ratings of the role of each of these factors in the evaluation of students varied depending on both grade level and subject area. Grade-level differences among elementary, junior high and high school teachers were dramatic. Only modest differences were found, however, in the factors teachers use in determining grades in science, social studies and language arts. The one exception in these subject area differences was teachers' ratings of the role of laboratories, which teachers reported using almost exclusively in science classes.

In work exploring the purpose of grades, Guskey (1996) further distinguished categories of evidence, identifying teachers' use of 'product' indicators of achievement that reflect what students currently know and are able to do; from 'process' behaviours that enable or support learning, such as homework and class participation; and 'progress' evidence that describes how much students have gained or improved. The distinction between 'product' and 'progress' makes it possible, for example, for students to demonstrate remarkable improvement but still receive low grades because their current level of achievement remains below grade level.

McMillan (2001) similarly differentiated evidence on students' academic achievement from teachers' consideration of 'academic enablers' (p. 25); such as effort, work habits, attention and participation; and other 'personal factors' related to students' personality and classroom behaviour. He found 'enablers' were more important to teachers in assigning grades than were 'personal factors,' a result that has been replicated in other studies (Duncan & Noonan, 2007; McMillan et al., 2002). Randall and Engelhard (2010) likewise found that teachers' beliefs about what behaviours best support learning are important in grading, especially when determining borderline grades.

Some researchers contend this multidimensional structure of grades stems from teachers' belief that 'academic achievement' conceptually includes behaviours that support and promote learning, especially at the elementary level (Chen & Bonner, 2016). The weight attached to these behaviours in determining students' grades has been shown to vary greatly among teachers, however, even those with similar teaching assignments within the same school (Cross & Frary, 1999; Duncan & Noonan, 2007; Guskey, 2009b; Webster, 2011).

Other researchers suggest these varying grading practices result at least in part from the lack of formal training teachers receive on grading and reporting (Stiggins, 2002). Most teachers have scant knowledge of the various grading methods or the effects of different grading policies on students (Brookhart & Nitko, 2008; Stiggins, 1993, 1999, 2008). Rarely do their grading policies and practices reflect those recommended by researchers or aligned with standards-based approaches (Guskey & Bailey, 2010, O'Connor, 2009; Reeves, 2011). It may be the case, however, that recent efforts to enhance teacher training programmes (e.g. Deluca & Klinger, 2010; Volante & Fazio, 2007) have resulted in newer teachers being more knowledgeable of effective

approaches to assessment and grading, and thus more likely to implement these practices.

This study was designed to address both of these issues by considering the influence of grade level and years of experience on teachers' use of both cognitive and non-cognitive factors in determining students' grades. Subject area differences were not considered because previous research has shown that with the exception of laboratory exercises, such differences are relatively modest or nonexistent (Gullickson, 1985), even among elementary teachers who typically teach multiple subjects (McMillan et al., 2002). In addition, by including a sample of more than 900 teachers, this investigation addressed a major limitation of most previous studies that have been conducted with relatively small, convenient samples of teachers.

Methods

Instruments

A questionnaire titled, *Teachers' Grading Practices Survey* (TGPS), was developed for use in the study. The TGPS consists of 17 selected-response items drawn from scales developed in the research of Liu, O'Connell, and McCoach (2006) and Guskey (2013a). Nine items ask teachers to record demographic data related to their current teaching assignment and experience. The other eight items address teachers' grading and reporting practices. Five of these items require yes/no responses, one requires a multiple-choice selection, one requires a 'check all that apply' selection, and one asks teachers to identify and attach a weight (up to 100%) to the elements they use in determining students' grades, such as major exams, reports, homework, class participation, etc. Teachers also can write in additional elements if none of those listed matches their grading approach.

Responses to this last item in the TGPS that provided the primary dependent variables in the study. It stated: 'Teachers use a variety of elements in determining students' grades. Among the elements listed below, please indicate those you use and about what percentage (%) each contributes to students' grades. Your selected total should add up to 100%.' The following elements were included:

- | | |
|--|--|
| <input type="checkbox"/> Major examinations | <input type="checkbox"/> Classroom observations |
| <input type="checkbox"/> Major compositions | <input type="checkbox"/> Oral presentations |
| <input type="checkbox"/> Unit tests | <input type="checkbox"/> Homework completion |
| <input type="checkbox"/> Class quizzes | <input type="checkbox"/> Homework quality |
| <input type="checkbox"/> Formative assessments | <input type="checkbox"/> Class participation |
| <input type="checkbox"/> Reports or projects | <input type="checkbox"/> Work habits and neatness |
| <input type="checkbox"/> Student portfolios | <input type="checkbox"/> Effort put forth |
| <input type="checkbox"/> Exhibits of students' work | <input type="checkbox"/> Punctuality of assignments |
| <input type="checkbox"/> Laboratory projects | <input type="checkbox"/> Class behaviour or attitude |
| <input type="checkbox"/> Students' notebooks or journals | <input type="checkbox"/> Progress made |
| <input type="checkbox"/> Other (Describe) _____ | |
| <input type="checkbox"/> Other (Describe) _____ | |

Most of these elements were derived from previous research studies on the sources of evidence and aspects of student performance teachers' use in determining grades (i.e. Brookhart, 1991, 1994; Gullickson, 1985; McMillan & Lawson, 2001; McMillan et al., 2002; Randall Engelhard, 2010). Others were added based on recent descriptions of teachers' grading practices (Guskey & Bailey, 2010; Reeves, 2011). Certain elements used in other research were purposefully excluded, however, due to their non-academic nature; for example, 'credit for bringing in items for a food drive' (McMillan, 2001, p. 24).

The majority of these elements can be classified as evidence of cognitive achievement. Major examinations and compositions, unit tests, class quizzes, reports or projects, portfolios, exhibits of students' work, laboratory projects and oral presentations would be included in this category. Several elements, however, could reflect either achievement or learning 'process' behaviours, depending on how the teacher uses each. Homework, for example, might include special projects completed primarily outside of class that demonstrate students' achievement of important academic goals. But homework also might be simply extended practise on concepts or skills students learned in class, much like attending rehearsals for a band or choral concert. Similarly, classroom observations could involve teachers recording students' performance of specific academic skills or simply noting if students actively contribute in group activities. Elements such as class participation, work habits and neatness, effort, punctuality of assignments and class behaviour or attitude would generally be classified as non-cognitive, 'process' behaviours that do not reflect achievement per se, but either enable learning or relate to classroom management (Bonnor & Chen, 2009).

It was recognised that some elements may overlap in teachers' interpretations. One teacher, for example, may use class quizzes as formative assessments. Another teacher may assemble major compositions into student portfolios. Pilot testing of the survey revealed, however, that teachers made distinctions between these when assigning weights to the various elements. Specifically, teachers who used class quizzes as formative assessments assigned weight for the grade to formative assessments and not to class quizzes. Likewise teachers who assembled compositions into student portfolios assigned weight to portfolios and not to individual compositions. In other words, they did not assign weight to both. So despite possible differences in teachers' interpretation of specific elements, teachers distinguished between elements in the weights they assigned.

Teachers were assured anonymity in their responses and only group data were reported.¹ Pilot testing of the TGPS grading items on a sample of 50 teachers (eliminating the weighted response item) showed it to have an internal reliability (α) of .87.

Data source

The survey was distributed via email to all teachers in five school districts in a Southeastern state in the U.S. during spring of 2016. These districts serve approximately 36,000 students within 49 schools and employ a total of 2233 full-time teachers. Three of the five districts are classified as suburban, with an average of 14.1% of the student population (14,043) coming from economically disadvantaged households. Two of the 13 suburban schools (15%) are classified as Title I schools; that is, schools with high numbers or high percentages of children from low-income families. One of the

participating five districts is classified as urban, with 78% of its 13,100 student population coming from economically disadvantaged households, and 24 of 26 of its schools (92%) classified as Title I. Lastly, one district is classified as urban/suburban with 35.3% of its 8506 students considered economically disadvantaged, and five of its ten schools (50%) classified as Title I schools.

A total of 943 of the 2233 teachers returned a fully completed TGPS for a response rate of 42%. All responses were recorded anonymously and no personal identifiers were included. The researchers secured proper permissions to conduct the survey research from the participating districts' administration and a university Institutional Review Board.

Results

The demographic characteristics of 943 responding teachers are shown in [Table 1](#). Teachers of varying years of experience were fairly evenly dispersed across grade levels. The largest group of teachers (38.1%) had 11–20 years of experience. The smallest portion of teachers (15.8%) taught at the middle school level. The correlation between years of experience and grade level was $r = .03$, indicating there is no linear relationship between these variables in this sample of teachers. In other words, teachers' years of experience was unrelated to the grade level at which they taught.

To determine if responding teachers differed systematically from non-respondents, comparisons were made between the demographic characteristics of responding teachers (i.e. teacher gender, grade level and district) and district averages. In all cases the proportions of teachers responding in each category were quite similar to the overall population of teachers in each district.

[Tables 2](#) and [3](#) show the means and standard deviations of the weights teachers assigned to each of these elements by grade level and by years of experience. As is evident in [Table 2](#), teachers at each grade level varied considerably in the weights they assigned to different elements. Nonetheless, differences are apparent in the weights assigned to several of the 22 elements. Specifically, teachers at the high school level (grade levels 9–12) appear to attach more weight to major compositions and examinations, laboratory projects and homework than do their colleagues at the elementary and middle school levels. Elementary teachers (grade levels K-5) appear to give more weight to formative assessments, exhibits of student work, and classroom observations.

Table 1. Percent and number of sample teachers by grade level and years of classroom experience.

Years of Classroom Experience	Grade Level				Total <i>n</i> = 943
	K-2 <i>n</i> = 228	3-5 <i>n</i> = 266	6-8 <i>n</i> = 149	9-12 <i>n</i> = 300	
0-3 <i>n</i> = 87	2.3% (22)	2.1% (20)	2.1% (20)	2.7% (25)	9.2%
4-10 <i>n</i> = 242	5.6% (53)	7.4% (70)	3.3% (31)	9.3% (88)	25.7%
11-20 <i>n</i> = 359	9.3% (88)	11.9% (112)	6.0% (57)	10.8% (102)	38.1%
21+ <i>n</i> = 255	6.9% (65)	6.8% (64)	4.3% (41)	9.0% (85)	27.0%
Total <i>n</i> <i>n</i> = 943	24.2%	28.2%	15.8%	31.8%	100.0%

Table 2. Means and standard deviations of the percentage weights teachers at different grade levels attach to different elements of student performance when determining grades.

Element	Grade Level			
	K-2 <i>n</i> = 228	3-5 <i>n</i> = 266	6-8 <i>n</i> = 149	9-12 <i>n</i> = 300
Major examinations	11.0 (17.1)	7.1 (12.3)	11.8 (16.6)	16.5 (18.4)
Major compositions	.6 (3.1)	2.2 (5.6)	1.8 (6.4)	4.2 (10.3)
Unit tests	14.5 (21.2)	15.8 (16.6)	19.3 (17.8)	18.7 (18.0)
Class quizzes	13.3 (19.3)	20.8 (18.4)	22.1 (14.1)	15.9 (12.2)
Formative assessment	24.5 (31.3)	16.2 (19.9)	7.3 (12.2)	5.3 (8.7)
Reports or projects	2.6 (5.7)	7.7 (8.1)	7.9 (10.0)	8.5 (11.1)
Student portfolios	1.8 (9.5)	.8 (3.8)	.6 (3.1)	1.6 (5.5)
Exhibits of students' work	5.4 (15.2)	4.1 (11.7)	1.7 (5.5)	1.7 (4.9)
Laboratory projects	.2 (1.1)	.9 (2.8)	2.1 (5.6)	3.9 (10.7)
Students' notebooks or journals	1.5 (3.6)	2.9 (4.6)	2.8 (5.8)	2.7 (5.8)
Classroom observation	6.2 (11.5)	2.8 (8.0)	1.8 (5.0)	1.6 (4.7)
Oral presentations	1.1 (7.1)	1.8 (3.8)	1.2 (3.3)	1.8 (4.8)
Homework completion	.5 (2.4)	1.3 (4.6)	4.5 (5.4)	4.5 (6.8)
Homework quality	.1 (6)	.6 (4.6)	1.7 (4.6)	1.9 (4.3)
Class participation	4.2 (11.2)	5.1 (12.7)	6.2 (15.1)	5.5 (8.6)
Work habits and neatness	.8 (3.3)	.4 (1.8)	.1 (7)	.3 (1.7)
Effort put forth	1.0 (3.1)	2.2 (8.8)	1.4 (5.1)	1.7 (7.7)
Punctuality of assignments	.2 (1.0)	.4 (1.9)	.5 (2.2)	.5 (1.6)
Class behaviour or attitude	1.0 (3.6)	.9 (7.2)	.2 (1.3)	.5 (2.1)
Progress made	1.7 (8.6)	.9 (6.1)	1.3 (7.4)	1.3 (7.1)
Other 1	11.0 (27.3)	5.0 (16.8)	4.5 (14.7)	3.0 (13.0)
Other 2	.7 (5.7)	.8 (5.3)	.9 (5.2)	.2 (1.9)

Note. Non-cognitive elements are identified in **bold print**.

Overall, about 10–20% of the weight used in determining students' grades is derived from non-cognitive factors such as class participation, work habits and neatness, effort, punctuality in turning in assignments, etc. (see all Non-cognitive elements identified in Table 2). Although this may appear to be a modest proportion, in a percentage grading system linked to letter grades; where a grade of 'A' is assigned to scores of 90–100%, a grade of 'B' to scores of 80–89%, and a grade of 'C' to scores of 70–79%, etc.; this can impact students' overall course grades (see Guskey, 2013b). There appear to be few differences in the overall weight attached to cognitive versus non-cognitive factors among teachers at the different grade levels.

The data in Table 3 show similar large variation among teachers with comparable levels of experience. However, there appears to be little difference in the weights assigned to the elements based on teachers' years of experience. Although the weights across elements differ, teachers assign nearly equivalent weights to the elements across all levels of experience.

To determine the degree of linear relationship of grade level and years of experience to the weights assigned to the different elements, correlation coefficients were computed. These are illustrated in Table 4. Because of the relatively large sample size, a conservative level of statistical significance of $\alpha < .001$ was selected for these tests.

Results showed that for 10 of the 22 elements, correlations between the weights teachers assigned and grade level were statistically significant. Specifically, as grade level increases so does the weight teachers assign to major exams and compositions, unit tests, homework completion, and homework quality. In addition, at lower grade levels teachers generally attach more weight to formative assessments, exhibits of students'

Table 3. Means and standard deviations of the percentage weights teachers with different years of experience attach to different elements of student performance when determining grades.

Element	Years of Experience			
	0–3 <i>n</i> = 87	4–10 <i>n</i> = 242	11–20 <i>n</i> = 359	21+ <i>n</i> = 255
Major examinations	13.6 (20.5)	11.0 (17.1)	10.1 (15.8)	10.4 (17.4)
Major compositions	1.2 (4.1)	2.7 (7.2)	2.6 (7.9)	1.9 (5.9)
Unit tests	18.6 (19.0)	18.3 (19.5)	16.3 (18.5)	14.1 (17.0)
Class quizzes	15.5 (13.3)	16.3 (15.8)	16.7 (15.9)	17.9 (18.2)
Formative assessment	16.8 (25.3)	11.3 (20.1)	11.2 (19.4)	15.9 (24.1)
Reports or projects	6.6 (9.7)	6.3 (8.1)	6.8 (9.4)	6.7 (10.6)
Student portfolios	1.0 (4.0)	1.8 (5.9)	1.5 (7.6)	1.3 (5.6)
Exhibits of students' work	1.8 (4.8)	3.6 (11.1)	3.5 (10.6)	3.2 (10.8)
Laboratory projects	1.6 (5.5)	1.8 (6.5)	1.9 (7.1)	1.9 (7.3)
Students' notebooks or journals	1.5 (3.4)	2.2 (4.8)	2.5 (5.1)	2.6 (5.5)
Classroom observation	2.9 (8.4)	3.2 (7.8)	3.7 (9.4)	3.8 (9.5)
Oral presentations	1.0 (2.5)	2.0 (7.1)	1.3 (3.4)	1.3 (4.6)
Homework completion	3.8 (6.9)	2.4 (5.3)	2.1 (4.8)	3.1 (5.8)
Homework quality	.8 (2.4)	1.0 (3.2)	1.1 (4.6)	.9 (3.0)
Class participation	4.2 (7.4)	5.0 (10.2)	5.9 (13.9)	6.1 (13.3)
Work habits and neatness	.6 (2.1)	.5 (2.5)	.4 (2.2)	.2 (1.5)
Effort put forth	1.2 (3.5)	2.6 (9.9)	1.5 (6.5)	1.5 (6.1)
Punctuality of assignments	.6 (2.0)	.5 (2.0)	.4 (1.2)	.1 (1.0)
Class behaviour or attitude	2.0 (11.2)	.8 (3.7)	1.2 (6.7)	.7 (4.0)
Progress made	1.2 (10.4)	1.2 (4.0)	1.4 (7.1)	1.2 (6.8)
Other 1	2.4 (13.0)	7.0 (19.5)	6.5 (19.8)	5.8 (18.8)
Other 2	1.0 (8.0)	1.2 (8.1)	.3 (3.2)	.5 (3.8)

work, classroom observations, and other sources of evidence. The most frequently noted other evidence sources were daily work, classwork, and class preparation (e.g. coming to class prepared and bringing essential supplies and materials). Similar to the results comparing means and standard deviations, none of the correlations between the weights teachers assigned to different elements and years of experience was statistically significant. Although teachers differ in the weights they attach to these different elements in determining students' grades, these differences appear to be unrelated to teachers' years of experience.

To test the statistical significance of these relationships, a multivariate regression analysis was conducted in which grade level and years of experience were considered the two independent variables, and weights attached to the 20 different elements of grading (eliminating the two 'Other' categories) the dependent variables. Again because of the large sample size, the more conservative level of $\alpha < .001$ was used for all tests of statistical significance. The results of this analysis, shown in Table 5, confirm that grade level contributed significantly to differences in the weights teachers assigned to these grading elements but years of experience did not. The interaction of grade level and years of experience also was not statistically significant.

Follow-up tests of between subject effects revealed statistically significant differences in 11 of the 20 dependent variables. These tests, along with calculated effect sizes (*d*), are displayed in Table 6. As had been shown in the analysis of correlations, teachers at upper grade levels put more emphasis on major examinations and compositions, unit tests, class quizzes, reports or projects, student portfolios, laboratory projects, and homework in determining students' grades. Teachers in lower elementary grades attach

Table 4. Correlation coefficients of grade level and years of experience to the percentage weights teachers attach to different elements of student performance when determining grades.

Element	Correlation Coefficients ($n = 943$)	
	Grade Level	Years of Experience
Major examinations	.26*	-.04
Major compositions	.18*	-.01
Unit tests	.14*	-.09
Class quizzes	.05	.06
Formative assessment	-.37*	.04
Reports or projects	.19*	.01
Student portfolios	-.01	.01
Exhibits of students' work	-.15*	.01
Laboratory projects	.21*	.05
Students' notebooks or journals	.06	.03
Classroom observation	-.20*	-.02
Oral presentations	.03	-.01
Homework completion	.30*	.01
Homework quality	.17*	.04
Class participation	.02	-.05
Work habits and neatness	-.05	-.04
Effort put forth	.02	-.11
Punctuality of assignments	.07	-.02
Class behaviour or attitude	-.07	-.01
Progress made	-.01	.02
Other 1	-.15*	-.05
Other 2	-.04	.02

* $p < .001$.**Table 5.** Multivariate tests for grade level and years of experience on the percentage weights teachers attach to different elements of student performance when determining grades.

Effect	Multivariate Tests ^a ($n = 943$)				
	Wilks' Lambda	F	Hypothesis df	Error df	Sig.
Intercept	.097	396.85	20.00	850.00	.000
Grade Level	.500	8.06	80.00	3355.56	.000
Years of Experience	.859	1.32	100.00	4151.41	.020
Grade Level x Experience	.595	1.13	400.00	12121.85	.038

^a Design: Intercept + Level + Experience + Level x Experience

more weight to formative assessment results, exhibits of student work and classroom observations of students.

Discussion

This research aimed to explore the factors teachers consider when determining the grades they record on students' report cards and if these factors vary by grade level and teachers' years of experience. Results of this study verify that at all grade levels, teachers use evidence of student learning as the primary factor in determining students' grades. Nevertheless, the specific sources of evidence teachers use were found to vary depending on the grade level of students.

As expected, secondary teachers at the middle and high school level tend to attach more weight to major examinations and compositions, unit tests, class quizzes, reports or

Table 6. Between subject tests (univariate) for grade level on the percentage weights teachers attach to different elements of student performance when determining grades.

Tests of Between-Subjects Effects (n=943)							
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	<i>d</i>
Grade Level	Major examinations	17,907.66	4	4476.91	16.84	.000	.267
	Major compositions	1375.54	4	343.88	7.52	.000	.179
	Unit tests	14,403.03	4	3600.75	11.04	.000	.216
	Class quizzes	17,910.36	4	4477.59	16.86	.000	.267
	Formative assessment	67,647.34	4	16911.83	45.58	.000	.440
	Reports or projects	2491.76	4	622.94	7.60	.000	.180
	Student portfolios	819.71	4	204.92	4.91	.001	.144
	Exhibits of students' work	1245.41	4	311.35	2.94	.020	.112
	Laboratory projects	1956.67	4	489.16	10.98	.000	.216
	Students' notebooks or journals	175.87	4	43.96	2.09	.080	.094
	Classroom observation	5870.45	4	1467.62	20.34	.000	.294
	Oral presentations	136.24	4	34.06	1.39	.234	.077
	Homework completion	2208.14	4	552.03	21.41	.000	.301
	Homework quality	319.66	4	79.91	5.43	.000	.152
	Class participation	188.62	4	47.15	.28	.890	.034
	Work habits and neatness	18.36	4	4.58	1.16	.327	.070
	Effort put forth	140.06	4	35.01	.58	.671	.047
	Punctuality of assignments	29.05	4	7.26	2.74	.027	.108
	Class behaviour or attitude	370.60	4	92.64	2.23	.064	.097
	Progress made	419.62	4	104.90	1.79	.128	.087

projects, student portfolios, laboratory projects and homework in determining students' grades than do elementary teachers (Link, 2018; Marso, 1985; Popham, 2009). The advanced language arts skills of middle and high school students allow secondary teachers to take advantage of assessment formats such as these that require reading and writing competencies many elementary students may not possess. The greater emphasis on homework may stem from secondary teachers' perception that older students should be taking greater responsibility for their own learning by completing certain learning tasks outside of class (Ellerbrock, Abbas, & DiCicco, 2016). Elementary teachers were found to attach more weight to formative assessment results, exhibits of student work and classroom observations of students than do secondary teachers. Depending on the nature of the formative assessments, these sources of evidence may require only modest language arts skills on the part of students, which would make them more appropriate at the elementary than secondary level.

Part of the reason for these grade level differences is undoubtedly due to the age-related appropriateness and validity of these various sources of evidence of student learning. Because younger students have limited writing skills, elementary teachers are less likely to use compositions, reports, and lab projects as ways to gather evidence on what students have learned. Instead, these early grade teachers rely more on students' exhibits or demonstrations of learning, along with their observations of students' performance in class.

Similarly, because younger students take part in fewer forms of summative assessment, elementary teachers may rely more on formative assessment results in determining grades. This is somewhat concerning, since formative assessments should be designed primarily to offer pertinent feedback on learning progress to both students and teachers and to guide the correction of learning difficulties, rather than to judge students' culminating performance and determine grades (Black & Wiliam, 2009). As

students move into more advanced grades and develop better writing skills, however, other forms of written expression and a broader range of assessment formats become appropriate and offer valid evidence of learning. Because students will likely experience a shift in teachers' expectations regarding assignments as they move from elementary to middle school (Link, 2018; Williamson, 2009), it will be important for teachers to provide additional support for students in this transition so they are well informed and familiar with the different sources of evidence that will be used to evaluate their learning.

An interesting unexpected result was the apparent non-linear relationship of grade level and several sources of evidence. Class quizzes, for example, were found to be given more weight by middle grade teachers than by early elementary teachers or by high school teachers. This finding is different from previous studies that found quizzes to be used more by both middle and high school teachers than teachers at the elementary level (Marso, 1985, Marso & Pigge, 1987; Mertler, 1999). Similarly, in this study student portfolios were emphasised more by the early elementary and high school teachers than by teachers in the middle grades, whereas earlier investigations reported elementary teachers using portfolios more frequently than either middle school or high school teachers (Marso, 1985, 1987; Mertler, 1999; Stiggins & Bridgeford, 1985). Perhaps as described earlier, early elementary teachers may not use quizzes as much due to the limited writing skills of their students, while high school teachers simply use a broader array of assessment instruments and need to rely less on classroom quiz results.

Additionally, the sources of evidence teachers use in determining students' grades were found not to vary depending teachers' years of classroom experience. We found no evidence that newly trained teachers with fewer years of experience weigh various sources of evidence of student learning any differently in determining students' grades than do their more experienced colleagues. This contrasts with the results of previous studies that showed teachers' classroom experience influences their approaches to assessment and grading (e.g. DeLuca, Valiquette, Coombs, LaPointe-McEwan, & Luhanga, 2016; Guskey, 2009a; Kauffman, Johnson, Kardos, Liu, & Peske, 2002; Mertler, 2003, 2004).

In contrast to their more experienced colleagues, newer teachers are more likely to complete pre-service teacher education programmes with a more contemporary understanding of effective assessment practices. Specifically, they are more likely to know about balanced assessment approaches that emphasise the integration of assessments *for*, *of* and *as* learning throughout classroom instruction. In addition, more recently trained teachers are also likely to have learned about modern standards- or competency-based approaches to grading that emphasise grades based on specific learning criteria so they have direct meaning and serve the communicative purposes for which they are intended (Chappuis & Stiggins, 2017; Deluca et al., 2017; Guskey, 2009a; Klinger, Volante, & DeLuca, 2012; Link, 2018; Popham, 2009; Stiggins, 1991).

As a result of their experiences in contemporary teacher education programmes, it was expected that newer teachers would choose different assessment methods as well as different sources of student learning evidence in determining students' grades than teachers trained at earlier times (Brookhart et al., 2016; Campbell, Murphy, & Holt, 2002; Plake, 1993). Lack of more up-to-date training may cause more experienced teachers to take more traditional approaches to assessment and grading in which they

consider not only evidence of student achievement but evidence of different ‘process’ variables such as homework, formative assessments, class participation, etc. in determining students’ grades (Guskey, 2015). This combination of student achievement and process variables may produce ‘score pollution,’ in which students’ grades do not represent academic mastery and limit ‘students, families and other stakeholders in the educational system from attaining valid information regarding academic achievement’ (Green, Johnson, Kim, & Pope, 2007, p. 1002).

These lack of differences in grading practices of teachers with varying years of classroom experience also may be a result of the high accountability climate evident across many schools in the United States today (Kauffman et al., 2002; Pizmony-Levy & Woolsey, 2017). Although newer teachers may have a more contemporary understanding of effective assessment and grading practices upon entry into the profession, they often shift to more standardised approaches to meet expected accountability demands (Deluca et al., 2017). Despite developments in standards-based assessment and grading reforms, new teachers are receiving little on-the-job guidance about how to teach and assess standards aligned to high-stakes testing and evaluations (Grainger & Adie, 2014; Pizmony-Levy & Woolsey, 2017). Newer teachers also may be complying with the pre-established grading norms of their more experienced colleagues or prescribed grading policies within their school or district, thereby fostering consistency in teachers’ grading practices over time.

Another possibility is that newer teachers may feel overwhelmed by the instructional demands of the classroom and, as a result, deprioritise grading. Instead of initiating new, more contemporary approaches to grading, they simply replicate whatever their more experienced grade-level or subject-specific colleagues have in place (Britt, 1997; Flores, 2006; Lawrence, Celis, & Ott, 2014). In other words, teaching context may be a contributing and even neutralising factor with regard to teachers’ grading practices.

Of course, grading similarities also could mean that recent improvements in teacher education programmes, especially those designed to help teachers develop greater assessment literacy and sounder grading and reporting practices, simply have had little impact on newer teachers’ assessment and grading practices. Perhaps despite the curricular changes in teacher education programmes, the assessment and grading practices of instructors and faculty members in those programmes remain quite traditional in nature (Alm & Colnerud, 2015; National Council on Teacher Quality, 2014). As a result, pre-service teachers may be learning about sound assessment and grading practices but not personally experiencing those practices. Lacking any personal experience or understanding of the potential impact of these practices, they may see little value in using them when they enter their professional teaching positions.

The results of this study further verify that in addition to using evidence of students’ learning to determine grades, teachers across all grade levels also use evidence based on non-cognitive factors typically related to aspects of students’ behaviour. The particular non-cognitive factors teachers consider vary depending on the grade level of the students. Early studies by Marso (1985), and Marso and Pigge (1987) showed that elementary teachers weight non-cognitive factors, such as ‘work habits and neatness’ and ‘class behaviour and attitude’ to determine grades more so than middle and high school teachers. Other research by Cizek, Rachor, and Fitzgerald (1995) and Liu (2008), however, did not report such difference in assessment and grading practices by grade level.

Contrary to Brookhart et al. (2016) century of research meta-analysis that found student effort to be a ‘key element in grading’ (p. 22), in this study teachers at all grade levels put little emphasis on ‘effort’ in determining students’ grades. Early elementary teachers (grades K-2) in this study weighed ‘effort put forth’ less, however, than their upper elementary (grades 3–5), middle, and secondary colleagues. Because early elementary curriculums and report cards typically include more non-cognitive elements, such as citizenship, participation, effort, etc., it was expected that teachers in the early grades would place greater emphasis on ‘effort’ and other non-cognitive factors when determining students’ grades (Guskey & Bailey, 2010; Guskey, Swan & Jung, 2011).

Limitations

Two design issues limit the strength of implications that can be drawn from this study. The first is the survey’s self-reported weightings. Without investigations into teachers’ gradebooks, it is difficult to determine if the weights teachers report on the survey correspond to their actual grading practices and are not affected by notions of social desirability in responses.

In addition, comparisons on demographic variables (i.e. teacher gender, grade level and district) showed the proportions of teachers responding in each category were quite similar to the overall population of teachers in each district. Still, we do not know if the teachers who responded systematically differed from those who chose not to respond in other critical variables. A higher response rate would improve the validity of results and allow greater certainty in the inferences drawn.

In addition, the survey instrument used in this research, the TGPS, asked teachers to consider only ‘what’ questions with regard to their grading practices. It did not probe further and ask ‘why?’ Hence, explanations of the reasons behind the identified differences in teachers’ responses are only speculative at this time. It would be helpful to know, for example, explicitly why teachers at the elementary level give more weight to formative assessment results in deciding students’ report card grades than do middle school and high school level teachers. Further research inquiring about the specific reasons why teachers at different grade levels use different sources of evidence in determining students’ grades will be helpful in clarifying these important issues.

Future directions

This study has implications for teacher professional development, pre-service teacher education programmes and researchers. Knowing the elements teachers use in determining students’ grades vary across grade levels can help target teacher professional development efforts aimed at improving teachers’ assessment and grading literacy. Recognising, for example, that both elementary and high school teachers rely more on student portfolios to show evidence of student learning than do middle school teachers could lead to professional development programmes specifically designed to help middle school teachers recognise the potential benefits of portfolio-based assessments. Similarly, knowing elementary teachers rely less on compositions as evidence of student learning than do middle and high school teachers could prompt professional

development initiatives focused on writing at the elementary level, especially upper elementary (grades 3–5). Moreover, improving assessment alignment across grades, adapted to the developmental skills of students, might ease students' transition from one grade level to the next, especially as students shift from elementary to middle school.

This study's results also show that in determining students' report card grades, elementary teachers rely more on students' exhibits or demonstrations of learning, along with observations of students' performance in class, than do middle and high school teachers. This shift in assessment evidence may be challenging for middle school students as they adjust to grading practices that are different from what they experienced in elementary school. Middle school students unfamiliar with summative assessments being used to communicate their achievement also may find it difficult transitioning from the more formative assessment emphasis of their elementary teachers. Teacher education preparation programmes may find these results useful in guiding improvements as well, especially in efforts to help new teachers align with contemporary assessment literacy and effective grading practices research.

In addition, assessment and grading researchers can help further explain the variability found across grade levels by studying the reasons why teachers prioritise particular types of assessment evidence at different grade levels as a means to communicate student learning. Researchers also may build on this study's findings by conducting longitudinal studies, following new teachers through their early years in the classroom, to help identify if teachers' assessment and grading practices change or evolve and, if so, the reasons for such change. Focus groups and in depth interviews of verbal protocols would be particularly useful in providing insights into different weighting strategies and to the reasons behind teachers' grading practices. Concerted efforts aligning the work of teacher educators, professional development leaders and researchers will positively contribute to the more thorough examination and improved implementation of effective grading policies and practices across all grade levels.

Educators throughout the world today struggle in their efforts to improve the way teachers grade and report on student learning progress in school. Most recognise that teachers vary widely in the procedures they use to determine students' grades and that the grades teachers assign typically are based on a 'hodgepodge' of different sources of evidence. They also acknowledge, however, that if teachers' classroom assessment practices yielded reliable and valid results, then many aspects of that evidence could represent accurate portrayals of what students have learned and are able to do. Grades based on such evidence can potentially provide the basis for appropriate and meaningful communication from teachers and schools to students and their families.

Note

1. All data analyses were conducted using IBM SPSS Statistics, Version 24. No alteration or customisation of the analysis techniques was employed.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Thomas R. Guskey is Professor Emeritus in the College of Education, University of Kentucky, Lexington, and he is the author/editor of 24 award-winning books and more than 250 book chapters, articles, and professional papers on educational measurement, evaluation, assessment, grading, and professional learning. He may be contacted at guskey@uky.edu, through Twitter at @tguskey, or at www.tguskey.com

Laura J. Link, Ed.D., is an Assistant Professor of Educational Leadership in the College of Professional Studies at Purdue University Fort Wayne. She is the co-author of *Cornerstones of Strong Schools: Practices for Purposeful Leadership* and author of several articles and professional papers on school leaders, grading, and assessments. Laura also leads K-16 research-practice partnerships and has twice won her college's Community Engagement Award.

Laura served in many K-12 central office and school-based leadership roles and has taught elementary, middle, high school and college students. While Laura was Assistant Superintendent of Teaching & Learning in Memphis, TN, she was 1 of 7 administrators charged with leading the largest school district merger in United States' history. Follow Laura on Twitter @laurajlink or contact her at linkl@pfw.edu

References

- Adrian, C. A. (2012). *Implementing standards-based grading: Elementary teachers' beliefs, practices and concerns* (Doctoral dissertation). Retrieved from ProQuest. (1032540669)
- Alm, F., & Colnerud, G. (2015). Teachers experiences of unfair grading. *Educational Assessment*, 20(2), 132–150.
- Bailey, M. T. (2012). *The relationship between secondary school teacher perceptions of grading practices and secondary school teacher perceptions of student motivation* (Doctoral dissertation). Retrieved from ProQuest. (1011481355)
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Bonner, S. M., & Chen, P. P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, 14, 57–77.
- Britt, P. M. (1997, November). *Perceptions of beginning teachers: Novice teachers reflect upon their beginning experiences*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Memphis, TN.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30, 123–142.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7, 279–301.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... 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., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson Education.
- Campbell, C., Murphy, J. A., & Holt, J. (2002 October). *Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Chappuis, J., & Stiggins, R. J. (2017). *An introduction to student involved assessment for learning*, 7th. New York, NY: Pearson.
- Chen, P. P., & Bonner, S. M. (2016, April). *Inservice teachers' grading beliefs*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

- Cizek, G. J., Rachor, R. E., & Fitzgerald, S. (1995, April). *Further investigations of teachers' assessment practices*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12, 53–72.
- Deluca, C., Coombs, A., LaPointe, D., & Chalas, A. (2017, April). *Changing approaches to classroom assessment: An empirical study across teacher career stages*. Paper presented at the annual meeting of the American Educational Research Association, San Antonio, TX.
- Deluca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438.
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 1–21, online.
- Duncan, R. C., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, 53, 1–21.
- Ellerbrock, C. R., Abbas, B., & DiCicco, M. (2016). Developmentally responsive teacher practices across the middle to high school transition. *Journal of Research in Education*, 24(1), 17–37.
- Flores, M. (2006). Being a novice teacher in two different settings: Struggles, continuities and discontinuities. *Teachers College Record*, 108(10), 2021–2052.
- Gareis, C. R., & Grant, L. W. (2015). Assessment literacy for teacher candidates: A focused approach. *Teacher Educators' Journal*, 2015, 4–21.
- Grainger, P. R., & Adie, L. (2014). How do preservice teacher education students move from novice to expert assessors? *Australian Journal of Teacher Education*, 39(7), 89–105.
- Green, S. K., Johnson, R. L., Kim, D., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23, 999–1011.
- Grimes, T. V. (2010). *Interpreting the meaning of grades: A descriptive analysis of middle school teachers' assessment and grading practices* (Doctoral dissertation). Retrieved from ProQuest. (305268025)
- Gullickson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research*, 79(2), 96–100.
- Guskey, T. R. (1996). Reporting on student learning: Lessons from the past – Prescriptions for the future. In T. R. Guskey (Ed.), *Communicating student learning. 1996 Yearbook of the association for supervision and curriculum development* (pp. 13–24). Alexandria, VA: Association for Supervision and Curriculum Development.
- Guskey, T. R. (2002, April). *Perspectives on grading and reporting: Differences among teachers, students, and parents*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Guskey, T. R. (2009a). Grading policies that work against standards... And how to fix them. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 9–26). Thousand Oaks, CA: Corwin.
- Guskey, T. R. (2009b, April). *Bound by tradition: Teachers' views of crucial grading and reporting issues*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Guskey, T. R. (2013a). Beyond tradition: Teachers' views of crucial grading and reporting issues. *Journal of Educational Research and Policy Studies*, 13(1), 32–49.
- Guskey, T. R. (2013b). The case against percentage grades. *Educational Leadership*, 71(1), 68–72.
- Guskey, T. R. (2015). *On your mark. Challenging the conventions of grading and reporting*. Bloomington, IN: Solution Tree Press.
- Guskey, T. R., & Bailey, J. M. (2010). *Developing standards based report cards*. Thousand Oaks, CA: Corwin.
- Guskey, T. R., Swan, G. M., & Jung, L. A. (2011). Grades that mean something: Kentucky develops standards-based report cards. *Phi Delta Kappan*, 93(2), 52–57.

- Imperial, P. (2011). *Grading and reporting purposes and practices in catholic secondary schools and grades' efficacy in accurately communicating student learning* (Doctoral dissertation). Retrieved from ProQuest. (896956719)
- Kauffman, D., Johnson, S. M., Kardos, S. M., Liu, E., & Peske, H. G. (2002). "Lost at sea": New teachers' experiences with curriculum and assessment. *Teachers College Record*, 104(2), 273–300.
- Klinger, D. A., Volante, L., & DeLuca, C. (2012). Building teacher capacity within the evolving assessment culture in Canadian education. *Policy Futures in Education*, 10(4), 447–460.
- Koloi-Keaikitse, S. (2012). *Classroom assessment practices: A survey of Botswana primary and secondary school teachers* (Unpublished doctoral dissertation). Ball State University, Muncie, IN.
- Lawrence, J. H., Celis, S., & Ott, M. (2014). Is the tenure process fair? What faculty think. *The Journal of Higher Education*, 85(2), 155–188.
- Link, L. J. (2018). Teachers' perceptions of grading practices: How pre-service training makes a difference. *Journal of Research in Education*, 28(1), 62–91.
- Liu, X. (2008, October). *Measuring teachers' perceptions of grading practices: Does school level make a difference?* Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Liu, X., O'Connell, A. A., & McCoach, D. B. (2006, April). *The initial validation of teachers' perceptions of grading practices*. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.
- Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues and Practice*, 27(3), 32–42.
- Marso, R. N. (1985, October). *Testing practices and test item preferences of classroom teachers*. Paper presented at the annual meeting of the Mid-Western Research Association, Chicago, IL.
- Marso, R. N., & Pigge, F. L. (1987, October). *Teacher-made tests: Testing practices, cognitive demands, and item construction*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20–32.
- McMillan, J. H., & Lawson, S. R. (2001). *Secondary science teachers' classroom assessment and grading practices*. Retrieved from ERIC database (ED 450 158).
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research*, 95, 203–213.
- McMillan, J. H., & Nash, S. (2000, April). *Teacher classroom assessment and grading decision making*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, New Orleans, LA.
- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120(2), 285–296.
- Mertler, C. A. (2003, October). *Pre-service versus inservice teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49–64.
- National Council on Teacher Quality. (2014). *Training our future teachers: Easy A's and what's behind them*. Washington, DC: Author.
- O'Connor, K. (2009). *How to grade for learning: Linking grades to standards* (3rd ed.). Glenview, IL: Pearson Professional Development.
- Pizmony-Levy, O., & Woolsey, A. (2017). Politics of education and teachers' support for high-stakes teacher accountability policies. *Education Policy Analysis Archives*, 25(87), 1–31.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment students. *Mid-Western Educational Researcher*, 6(1), 21–27.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental. *Theory Into Practice*, 48, 4–11.

- Randall, J., & Engelhard, G. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement*, 46, 1–18.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26, 1372–1380.
- Reeves, D. B. (2011). *Elements of grading*. Bloomington, IN: Solution Tree Press.
- Russell, J. A., & Austin, J. R. (2010). Assessment practices of secondary music teachers. *Journal of Research in Music Education*, 58, 37–54.
- Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and practice*, 10(1), 7–12.
- Stiggins, R. J. (1993). Teacher training in assessment: Overcoming the neglect. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 27–40). Lincoln, NE: Buros Institute of Mental Measurements.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23–27.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758–765.
- Stiggins, R. J. (2008). Report cards: Assessments for learning. In *Student-involved assessment for learning* (5th., pp. 267–310). Upper Saddle River, NJ: Merrill/Prentice.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271–286.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5–14.
- Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21, 326–343.
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings': An attempt to verbalize and discuss teachers' internalized grading criteria. *European Physical Education Review*, 20, 199–214.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770.
- Webster, K. L. (2011). *High school grading practices: Teacher leaders' reflections, insights, and recommendations* (Doctoral dissertation). Retrieved from ProQuest. (3498925)
- Williamson, R. (2009). *Scheduling to improve student learning*. Westerville, OH: National Middle School Association.
- Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, 116, 2998–3004.
- Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors* (Doctoral dissertation). Retrieved from ProQuest. (913076079)