

Examining School Principals' Conceptions of Assessment and Grading Practices

Liying Cheng, David Baidoo-Anu, Christopher DeLuca

Queen's University

Principals play a key leadership role in school effectiveness and student success; however, one area that has received relatively little attention so far is principals' embedded understanding of assessment and grading within the educational context where they work. We examined 141 Chinese school principals' conceptions of assessment and grading practices using a survey research design. Our results indicated that principals viewed summative assessment as necessarily serving two purposes: (a) accountability, and (b) improvement of teaching and learning. Assessment that does not serve these dual purposes was not viewed as maximally supporting learning in this educational context. The study further found that accountability pressures influenced principals' conceptions of assessment and consequently shaped their approaches to grading in China's examination-oriented culture. This study provides implications for a heightened understanding of principals' leadership in the crucial role that student assessment plays in both enhancing instruction and student learning and ever-expanding accountability mandates within the Chinese education context and other educational contexts globally.

Les directeurs d'école jouent un rôle clé dans l'efficacité de l'école et la réussite des élèves. Cependant, un domaine qui a reçu relativement peu d'attention jusqu'à présent est la compréhension de l'évaluation et de la notation par les directeurs dans le contexte éducatif où ils travaillent. À l'aide d'une enquête, nous avons examiné les conceptions de 141 directeurs d'école chinois sur les pratiques d'évaluation et de notation. Nos résultats indiquent que les directeurs d'école considèrent que l'évaluation sommative doit nécessairement servir deux objectifs : (a) la responsabilité et (b) l'amélioration de l'enseignement et de l'apprentissage. L'évaluation qui ne sert pas ces deux objectifs n'est pas considérée comme un soutien maximal à l'apprentissage dans ce contexte éducatif. L'étude a également révélé que les pressions liées à l'obligation de rendre des comptes ont influencé les conceptions des directeurs d'école en matière d'évaluation et, par conséquent, leurs approches de la notation dans la culture chinoise axée sur les examens. Cette étude a des implications pour une meilleure compréhension du leadership des directeurs d'école en ce qui concerne le rôle crucial que joue l'évaluation des élèves dans l'amélioration de l'enseignement et de l'apprentissage des élèves, ainsi que des mandats de responsabilisation toujours plus étendus dans le contexte éducatif chinois et d'autres contextes éducatifs dans le monde.

Principals' Roles in Assessment and Grading Within the Global Contexts

The role of principals in providing leadership in the area of assessment has only heightened in the COVID-era in educational contexts around the world. More than ever before, teachers, students,

and parents need sound guidance on how to navigate the shifting landscape of teaching, assessment, and school accountability. Current technological advancement has transformed the way in which assessment has occurred in physical and virtual classrooms, calling teachers to use new practices, methodologies, and technologies, and raising questions about what is valuable and what should be prioritized in teaching and learning moving forward (Doucet et al., 2020). Understanding how principals conceptualize assessment as a broad instructional practice and grading as a specific summative evaluative practice in relation to educational priorities globally is critical as school systems world-wide respond to the pandemic and begin to return back to a new normal (DeLuca et al., 2021).

Globally, school principals are considered educational leaders with the responsibility of ensuring school cultures promote student learning. According to the Alberta Teachers' Association, "School principals have the mandate of focusing on improving student learning and development through effective leadership practices" (Alberta Teachers' Association, 2004, p. 5). As part of school effectiveness and improvement frameworks, principals have a responsibility to align curriculum, instruction, and assessment in classrooms (Lunenburg, 2010; Marzano & Waters, 2010). Principals' roles are multivariate across many educational contexts. In particular, Stiggins and Duke (2008) perceive principals as assessment leaders with the mandate to support their teachers' assessment practices. Principals share an instrumental role in developing teachers' capacity to use contextually appropriate assessment practices to enhance effective teaching and learning (Ball, 2016). This role has become ever more prominent within the context of the global pandemic when teaching, learning, and assessment take on drastically different practices, methodologies, and technologies. Marzano et al. (2005) conducted a meta-analysis of 69 studies involving 2802 schools in the United States, approximately 1.4 million students, and 14,000 teachers. The analysis covered 35 years of research to explore what school leadership works, and concluded that principals' engagement in curriculum, instruction, and assessment has a positive relationship ($r = 0.20$) with student academic achievement. The study further found a positive relationship ($r = 0.25$) between principals' knowledge of curriculum, instruction, assessment, and students' academic achievement. Therefore, understanding the diverse ways that principals perceive assessment and how it connects meaningfully with grading practices in the current educational context is critical to both support principal and teacher professional development and students' consistent assessment experiences in schools, ultimately enhancing school effectiveness and student success.

Within individual educational contexts, assessment has increasingly become an integral component of effective teaching and learning, which mediates the interaction between teachers and students in the classroom. The ultimate goal of educational assessment is well-recognized globally to improve teaching and learning (Black & William, 2010; Brookhart, 2011; Ontario Ministry of Education, 2010, Ministry of Education of the People's Republic of China, 2013). Educational researchers have also consistently found a positive relationship between high-quality teaching, classroom assessment, and student learning (Black & William, 2006; Fenstermacher & Richardson, 2005; Hattie, 2009; Herppich et al., 2018; Putnam & Borko, 1997). However, Torrance (2001) found that educational policies are mostly oriented toward raising student academic achievement and making the school system more accountable.

In many educational contexts world-wide, such as China in particular, students do not graduate unless they pass the examinations. This function has been heightened during the global pandemic. In addition, principals' and teachers' performance are evaluated based on student achievement in standardized testing. Demand for accountability as reflected in high-stakes large-

scale testing programs continues to pose a significant risk to the uptake of authentic and sustained formative assessment practices in school systems (Birenbaum et al., 2015). In response to increasing accountability demands in schools, there has been further pressure on teachers to develop capacity to engage in high-quality classroom assessment and grading practices (Baidoo-Anu & Ennu Baidoo, 2022; Cheng et al., 2004; Cheng & Wang, 2007; Herppich et al., 2018; Popham, 2009; Xu & Brown, 2016). However, solely developing teachers' assessment literacy within the school system can result in an imbalance of assessment knowledges and interfere with professional relationships (Parker, 2006). It is imperative that principals are not left out of the assessment literacy equation as teachers and principals work together to support student learning in schools (Parker, 2006). This study aims to begin to address the lack of empirical literature on principals' assessment literacy, with a specific focus on Chinese principals' conceptions of assessment and grading practices.

Research Context

With nearly 260 million students and over 15 million teachers within approximately 514,000 schools, China is considered as the country with the largest education system in the world (National Bureau of Statistics of China, 2014). China's educational system is known to be highly dominated by high-stakes large-scale tests and examinations, notably at the end of middle school and senior secondary school (i.e., Zhong Kao 中考 and Gao Kao 高考 respectively; Brown & Gao, 2015; Cheng, 2008; He et al., 2011; Li, 2009; Liu & Qi, 2005; Niu, 2007). There is an over reliance on public examinations, consequently, students' performance in public examinations is used as a yardstick to gauge quality of teachers, schools (school principals), and the total worth of students (Brown & Gao, 2015; China Civilization Center, 2007). Recently, Emler et al. (2019) stated the major side effects of high-stake examinations in education including "distorting education, exacerbating inequity and injustice, demoralization of professionals, ethical corruption, and stifling of innovation in education" (p.1). Such side effects are particularly evident in the Chinese education system (Cheng, 2008; Zhao, 2014). The dominance of high-stakes examination system has introduced discords in Chinese teachers' assessment practices (Brown & Gao, 2015; Li & Hui, 2007) and has rendered assessment practices more summative than formative (Li, 2012). These high-stakes large-scale tests and examinations reduce formative assessment to a mini-summative assessment or to a series of teaching techniques for tests (Bennett, 2011; Hargreaves, 2013, Klenowski, 2009; Ma & Cheng, 2016; OECD, 2013).

Consequently, to reduce over reliance on high-stakes tests and examinations and improve formative assessment practices within the school system, the government of China in 2014 initiated reforms that aimed at establishing a contemporary examination system comprised of standardised tests, comprehensive evaluation, and assorted admissions criteria (OECD, 2016). Acknowledging the pivotal role school principals play in classroom assessment, school effectiveness, and student success, the Chinese government established a myriad of policies to enhance the quality of school principals. As part of the policies, principals are mandated to take over 360 class hours of training every five years to develop new knowledge and skills and improve on their managerial skills (Ministry of Education of the People's Republic of China, 2012). Most research on school principals thus far has largely focused on their role in many aspects of schooling (Li et al., 2012; Lo, 2004), however, one area that has received relatively little attention is Chinese principals' understanding of assessment and grading. Hence, the purpose of this study is to examine principals' conceptions of assessment and grading practices.

Given the contemporary roles of principals as instructional leaders with the core mandate to improve learning, encourage collaboration, and align curriculum, instruction, and assessment, it is necessary to study how they conceive assessment broadly and grading specifically within the school setting. Equally important is the understanding of how principals play such leadership roles within and across different educational settings and countries where instruction and assessment are socio-culturally situated and bound. Despite the significance and impact of assessment and grading on teaching and learning, researchers have long recognized the lack of theoretical grounding for teachers' assessment and grading practices. Specifically, researchers have called for an examination of assessment and grading practices using contemporary validity theories (Brookhart, 2013; Moss, 2003) which integrate a socio-culturally situated argument on the alignment of assessment and grading practices, values, and consequences (DeLuca, 2011; Moss, 2003). The socio-culturally situated theories argue for the understanding of embedded teaching, learning, and assessment culture in context (Chalhoub-Deville, 2003; Shepard, 2000), which provides a strong rationale for conducting research in one context with implications across educational systems.

The understanding of culture is not a new concept and has been studied in various fields beyond education (see Trompenaars & Hampden-Turner, 1998). The authors pointed out that "the essence of culture is not what is visible on the surface. It is the shared ways groups of people understand and interpret the world" (p.3). The shared views among teachers and principals determine how assessment and grading function in instruction. Kennedy (2007, 2016) used a socio-cultural framework to understand assessment practices in Asia and in Confucian-Heritage Cultures. Discussing the increasing importance of participating in international large-scale assessments, Kennedy raised the purposes of assessment and the cultural contexts in which they are embedded as a critical issue in Asia. For example, certain forms of assessment such as examinations have been privileged in some Asian countries, China in particular, in ways that they are not in many western countries (Morrison, 2006).

So far little is known about the underlying teaching and learning values that contribute specifically to grade decisions. For example, the concept of effort and behaviour as part of a grade is highly valued in Chinese learning culture (Carless, 2010; Cheng & Curtis, 2009) whereas grading only on achievement is more widely endorsed at the policy level in measurement-driven communities in North America (O'Connor, 2007; Simon et al., 2010; Wormeli, 2006). Our aim in this study is to examine Chinese principals' conceptions of assessment and grading practices, and the relationship between the two, within the examination-oriented culture in China.

Literature Review

Assessment Conceptions

Assessment conceptions refers to educators' overall perception and awareness of assessment (Barnes et al., 2017). Vandeyar and Killen (2007) emphasized that different assessment conceptions lead to different assessment practices. Assessment conceptions have been predominately grouped into educators' agreement or disagreement with four purposes for which assessment is used: (a) improvement of teaching and learning, (b) school accountability, (c) student accountability, and (d) treating assessment as irrelevant (Brown, 2004, 2006; Brown & Harris, 2009; Brown, Hui et al., 2011; Brown, Lake, & Matters, 2011; Brown & Remesal, 2012;). A similar study was conducted by Li and Hui (2007) that translated Brown's (2004) scale of

assessment conceptions (inventory TCOA-III) into Chinese to examine teachers' conceptions of assessment in China. The study grouped teachers' conceptions into three factors: negative, improvement, and accountability. Based on the findings, Li and Hui (2007) found that Chinese teachers had contradictory conceptions about assessment. These two researchers attributed their contradictory conceptions to the overemphasis on high-stakes large-scale examinations in China, rather than judgement of professionals as the dominant assessment policy and practice in other contexts. High-stakes large-scale examinations originated from China more than 2000 years ago and continue to dominate its educational assessment system (Cheng & Curtis, 2009; Paine, 1990, Zhao, 2014). These examinations are not only used in making critical decision about examinees such as job and educational selection and placement, but also determine self-worth or efficacy and virtue in Chinese society (Brown & Gao, 2015; Cheng, 2008; China Civilization Center, 2007). As mentioned earlier, the dominance of the high-stakes examination system has created tension in Chinese teachers' assessment practices (Brown & Gao, 2015; Li & Hui, 2007) and has rendered their assessment practice more summative than formative (Li, 2012). Although considerable studies (Brown & Gao, 2015; Li & Hui, 2007) have been conducted on Chinese teachers' conceptions of assessment, there is paucity of studies on Chinese principals' conception of assessment.

Principals have the responsibility to support teachers' classroom assessment practices (Lunenburg, 2010; Marzano & Waters, 2010). Principals need to understand features of assessment policies, set specific goals for the integration of assessment into instruction, and assist teachers in reaching those goals (Arter et al., 1993). Stiggins and Chappuis (2006) argued that, unfortunately, teachers do not have consistent support for classroom assessment practices as assessment training remains "virtually nonexistent in leadership training programs" (p. 11). Similarly, Arter et al. (1993) maintained that most principals are not trained in the basic principles and applications of assessment, either during their leadership training and certification programs or during their teaching preparation programs. As a result, the majority of school principals lack expertise in assessment. They are not able to support quality policy-level assessments and find it difficult to help teachers conduct classroom assessment effectively (Arter et al., 1993).

Grading

Grading is one of the most critical assessment practices that plays an important role in students' schooling and life. Grading is defined as the process of collecting and evaluating evidence of student achievement, performance, and learning skills and assigning grades for students on standardized report cards (DeLuca et al., 2019; Tierney et al., 2011). More importantly, there is a strong relationship between assessment and grading, as both practices incorporate students' achievement and behaviour (McMillan et al., 2002). Due to increasing school, teacher, and student accountability, grading has become one of the most high-stakes classroom assessment practices (Brookhart, 2013). Grades help provide information about students' achievement for stakeholders (Cheng & Sun, 2015). Grading increase students' self-esteem, determine students' eligibility for admissions into university and college programs, determine access to funding and awards, and are used for making other high-stakes decisions (DeLuca et al., 2019). In view of the above, grading has become one of the most vital assessment decisions teachers have to make in their instruction. Nitko (2001) maintained that in order not to distort the validity of grades, teachers should be extra careful when assigning grades because they are used by "students, parents, other teachers, guidance counsellors, school officials, postsecondary educational

institutions, and employers” (p. 365). Due to the high-stakes nature of grading, it is important to understand how principals, as school leaders, understand, support, and endorse grading practices. Such leadership could have tremendous impact on the nature and practices of assessment and grading in a school context.

Considering the grading practices in China, Liu and Yan (2016) found that grading policies are broadly defined and implicitly embedded across several government and non-government documents. In a comparative analysis of grading policies across Canada and China, Cheng et al., (2018) found that the “primary purpose of grading in China is to make a holistic evaluation of both the learner and his/her learning, instead of only assessing the learner’s academic performance” (p. 9). These researchers found that grading decisions are based on the process of learning and the outcomes of the learning process. Therefore, educators are encouraged to grade student achievements using both summative tests and formative assessments using diverse assessment methods. Cheng and Sun (2015) investigated 350 Chinese secondary school English language teachers’ grading decision making. The study found that teachers considered achievement and non-achievement factors in grading practices. Teachers placed even greater weight on non-achievement factors, such as effort, homework, and study habits. Teachers in this study also used multiple types of assessment in their grading practices.

Despite the consistent pressure on teachers to engage in quality grading practices, many educational researchers have argued that assigning grades to students is a difficult task for most teachers (Allen, 2005; Kohn 1999; Marzano 2000). Teachers often find difficulties in handling non-achievement factors, including students’ efforts, behaviours that ideally should not be included when grading (Allen, 2005; Kohn 1999; Marzano 2000; Randall & Engelhard, 2010). Moreover, teachers sometimes confuse the communication function of grades and therefore, attempt communicating multiple pieces of information about students which cannot be expressed in a single academic mark. If experienced teachers find grading a difficult task, new teachers will find grading extremely difficult and therefore may need more support (Allen, 2005; Kohn 1999; Marzano 2000). Given the difficulties teachers face in their grading practices, school principals are in the leadership position to support, supervise, and monitor teachers’ grading practices. Further, support from principals will not be possible or effective if principals themselves have low knowledge or do not understand contextually appropriate grading practices (Lentz, 1997). Despite the significant influence principals have on teachers’ grading practices, there is a paucity of studies, so far, on principals’ grading practices.

Gleaning from above, it is evident that principals play a key leadership role in school effectiveness and student success and in teachers’ assessment and grading practices. Previous research has explored this leadership role extensively in many aspects of schooling (Lunenburg, 2010; Marzano & Waters, 2010; Stiggins & Duke 2008). However, one area that has received relatively little attention is principals’ understanding of assessment and grading. In order to address this research gap, our study aimed to answer the following two questions.

1. What are Chinese principals’ conceptions of assessment and grading practices?
2. What is the relationship between these principals’ conceptions of assessment and grading practices?

Methods

The study examined Chinese school principals’ conceptions of assessment and grading practices

using a survey research design. Due to the global COVID-19 pandemic, the data collection was interrupted. A limited total of 141 Chinese principals (127 from around Suzhou Zhenjiang province and 13 from Shanxi province) responded to the survey, which consists of two scales: Conceptions of Assessment and Grading Practices. The study received ethical clearance from Queen's University General Research Ethics Board (GREB).

Participants

As Table 1 shows, out of the total participants of 141 school principals, 39.1% (45) were females and 53.2% (75) were males. A majority 84.4% (119) of the school principals had a master's degree. The level of experience of the participants were within the K–12 schooling context, with 59.6% in grades K–3. The school size of these principals ranged widely from 100–4305 students. All the principals stated that their students write large-scale examinations—at the school board (61.7%), provincial (31.2%), and national (12.8%) level.

Table 1

Demographics of the Participants

Demographic Variables	Frequency	Percentage
Gender		
Female	45	39.1
Male	75	53.2
I do not identify within the gender binary	8	5.7
Prefer not to respond	13	9.2
Educational Background		
Bachelor's Degree	2	4.3
Master's Degree	119	84.4
Associate Degree	13	9.2
Others	3	2.1
Level of Experience		
Junior Primary (grades K–3)	84	59.6
Senior Primary (grades 4–6)	37	26.2
Junior Secondary (grades 7–9)	19	13.5
Senior Secondary (grades 10–12)	1	.7
Participation of Large-scale Testing		
School board level	87	61.7
Provincial level	44	31.2
National level	18	12.8
Assessment Experience		
None	23	16.3
½ to 1-day Workshop or Seminar	6	4.3
Completed postgraduate course	44	31.2
Some hours as part of pre-service training	53	37.6
Completed undergraduate course	6	4.3
Completed professional courses	9	6.4

In relation to assessment training, 31.2% (44) of the participants had completed a postgraduate course in assessment, and another 37.6% (53) had some hours as part of pre-service training in assessment. 20.6% of the principals had no experience or 1/2 to 1-day workshop/seminar on assessment. The percentage of principals with relatively low assessment training is alarming, given the leadership role principals play in assessment practices in schools.

Instrument

A survey was used for this study that consisted of three sections: conceptions of assessment, grading practices, and participant demographic information. The Conceptions of Assessment Scale (Brown, 2006), which was composed of 27 items, was previously translated into Chinese. The items were measured using six-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Participants were also provided with an *I don't know* option for each item due to the nature of the items, which could have provoked an *I don't know* response. This scale was originally developed and validated in Queensland, Australia (Brown, 2006) and has been extensively used and validated in various cultural contexts including Canada (Daniels & Poth, 2017), Hong Kong (Brown et al., 2009), Cyprus (Brown & Michaelides, 2011), Egypt (Gebril & Brown, 2014), and India (Brown & Gao, 2015).

The grading practices scale consisted of 16 grading scenarios, which explored four dimensions of grading practices that had been previously identified in grading research: (a) classroom context, (b) learning values of grades, (c) policy and external pressures, and (d) consequences of grade use (Cheng et al., 2018). Following previous grading scales (e.g., Bonner & Chen, 2009; Cheng et al., 2020), the grading practices scale used 16 scenarios to provide the context for principals to elicit and interpret grading practices. Each dimension was also measured using a six-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Validity and reliability of the scale were considered. First, three researchers crafted the scenarios based on previous literature and a multi-dimensional conceptual framework (e.g., Bonner & Chen, 2009; Sun & Cheng, 2013). Then, an expert panel review consisting of five assessment experts was used to qualitatively examine the wording of the scenarios, and mapped the scenarios to the four underpinning dimensions (i.e., classroom context, learning values of grades, policy and external pressures, and consequences of grade use). Moreover, exploratory factor analysis and reliability analysis were conducted. Reliability analysis using Cronbach alpha showed sufficient internal consistency of 0.723. These sources provide evidence for valid interpretations of results from the grading scale.

Data Collection and Analysis Procedures

The surveys were administered to the participants via a Chinese on-line survey platform (问卷星 <https://www.wjx.cn/>) over a period of 6 months. Descriptive statistics (mean and standard deviation) were first calculated for all items to explore principals' conceptions of assessment and grading practices.

In order to understand the potential structural patterns emanating from principals' responses to conceptions of assessment and grading scenarios within a unique Chinese context, exploratory factor analysis was conducted. Before commencing the factor analysis, data were inspected to ensure suitability for factor analysis. The exploratory factor analysis was conducted in two stages—the conceptions of assessment and then the grading scenarios. In both stages, principal

axis factor analysis was employed because it is legitimate for short scale response options and small sample sizes (Bandalos & Finney, 2018). Oblique rotation (oblimin) was used to examine the component correlation matrix. In both stages one and two, exploratory factor analysis was conducted on 27 conceptions of assessment items and 16 grading practices scenario items using principal axis factor analysis separately. For both stages, the Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = .756, and .735 respectively (“middling” according to Hutcheson & Sofroniou, 1999). This means the sample was adequate for carrying out factor analysis. Bartlett’s test was significant (.000) for both stages, indicating that the assumption of sphericity was met.

Stage I Conceptions of Assessment

An initial analysis was run to obtain eigenvalues for each factor in the data. Eight factors had eigenvalues over Kaiser’s criterion of 1 explaining 60.3% of the variance. In order to determine the best factor structure pattern, Monte Carlo PCA for Parallel Analysis (Watkins, 2000) was used. The results of this analysis supported three factors, explaining 39.9% of the variance and with eigenvalues greater than the corresponding criterion values in a random data set, generated with a matrix of the same size (27 variables × 141 participants). Factor loadings below 0.4 were suppressed because they did not represent significant values (Field, 2013). Items 27, 31, 18, 24, 42, 17, 19, and 25 did not load onto any factor and were thus removed. Items that cluster on the same factor suggest that Factors 1, 2, and 3 represent assessment as summative, assessment as irrelevant, and assessment as improvement respectively. All three factors had an internal consistency (i.e., Cronbach’s alpha) greater than 0.70, indicating good reliability (Field, 2013). The three-factor structure was used to understand principals’ conceptions of assessment. Moreover, Pearson product moment correlation was used to understand the relationships between principals’ conceptions of assessment and grading practices based on the various factor structures found through the exploratory factor analysis.

Stage II Grading Practices

An initial analysis was run to obtain eigenvalues for each factor in the data. Five factors had eigenvalues over Kaiser’s criterion of 1 explaining 57.3% of the variance. In order to determine best factor structure pattern, Monte Carlo PCA for Parallel Analysis (Watkins, 2000) was used. The results supported two factors, explaining 33.72%. Factor loadings below 0.4 were suppressed because they did not represent significant values (Field, 2013). Items 1, 2, 3, 6, and 7 did not load onto any factor and were thus removed. The items that cluster on the same factor suggest that factor 1 represents consequential grading and factor 2 represents comprehensive grading. All factors had an internal consistency (i.e., Cronbach’s alpha) of greater than 0.70, indicating good reliability (Field, 2013). The two-factor structure was used to understand principals’ grading patterns.

Results

In order to understand school principals’ conceptions of assessment and grading practices, and the relationship between the two, descriptive statistics and factor analysis in assessment and grading are reported followed by a correlational analysis.

Descriptive Statistics

Table 2 shows the results of school principals' conceptions of assessment. Using a six-point Likert-type scale ranging from *not at all likely* (1) to *highly likely* (6), and with an additional *I don't know* item response (0), principals were asked to indicate their agreement or disagreement with the items presented. *I don't know* responses were coded as 0. Principals' levels of support for each item were calculated by averaging the priority they had assigned across the items presented in the

Table 2

Conceptions of Assessment Descriptive Statistics

Statements	Mean	SD	Skewness	Kurtosis
<i>To what extent do you agree or disagree with the statements below?</i>				
Q17 Assessment provides information on how well schools are doing.	4.57	1.18	-1.53	2.55
Q18 Assessment places students into categories.	2.70	1.28	.69	.10
Q19 Assessment helps determine how much students learn from teaching.	4.27	1.05	-.59	.39
Q20 Assessment provides feedback to students about their performance.	4.94	.98	-1.71	5.49
Q21 Assessment is integrated with teaching practice.	5.14	.91	-1.85	5.75
Q22 Assessment results are trustworthy.	4.28	.72	-.11	.84
Q23 Assessment forces teachers to teach in a way against their beliefs.	2.52	1.1	1.64	4.94
Q24 Teachers conduct assessments but make little use of the results.	3.40	1.18	.25	.25
Q25 Assessment results should be treated cautiously due to measurement error.	4.77	1.03	-1.12	1.92
Q26 Assessment is an accurate indicator of a school's quality.	3.62	.98	-.21	.39
Q27 Assessment is assigning a grade or level to student work.	3.33	1.20	.46	.42
Q28 Assessment establishes what students have learned.	3.57	1.12	-.09	.61
Q29 Assessment provides feedback to students about their learning needs.	4.38	.97	-.78	2.46
Q30 Assessment information modifies ongoing teaching of students.	4.76	1.04	-1.69	4.13
Q31 Assessment results are consistent.	2.88	.92	.90	3.02
Q32 Assessment is unfair to students.	2.45	1.05	2.02	6.90
Q33 Assessment results are filed and ignored.	2.47	1.12	1.60	4.43
Q34 Teachers should take into account the error in all assessment.	5.03	.92	-1.80	6.60
Q35 Assessment is a good way to evaluate a school.	4.01	.93	-.23	.30
Q36 Assessment determines if students meet curriculum expectations.	4.40	.77	.09	.65
Q37 Assessment measures students' higher order thinking skills.	3.57	1.04	.32	.75
Q38 Assessment helps students improve their learning.	4.74	.87	-.32	.71
Q39 Assessment allows different students to get different instruction.	4.67	.90	-.24	.97
Q40 Assessment results can be depended on.	4.06	.92	-.22	2.78
Q41 Assessment interferes with teaching.	2.52	1.07	1.39	4.66
Q42 Assessment has little impact on teaching.	3.52	1.25	.11	-.33
Q43 Assessment is an imprecise process.	3.19	1.28	.76	.96

Conceptions of Assessment scale. A mean of 3.50 and above indicates principals' agreement with the item and a mean below 3.50 indicates principals' disagreement with the item. The overall descriptive statistics from Table 2 show principals' agreement to most of the items on the Conceptions of Assessment scale. As a key finding, principals did not perceive assessment as independent of teaching. They believed assessment is highly integrated with teaching and learning (Item 21; $M = 5.14$, $SD = .91$). Notwithstanding, principals perceived assessment as having little impact on teaching (Item 42; $M = 3.52$, $SD = 1.25$). Moreover, principals agreed that assessment is an effective way of improving students' learning (Item 38; $M = 4.74$, $SD = .87$). Overall descriptive statistics showed that principals have positive views of assessment.

Table 3 shows the results of school principals' grading practices. Using a six-point Likert-type scale (ranging from *not at all likely* (1) to *highly likely* (6)), principals were asked to indicate their agreement or disagreement with 16 scenarios presented. Principals' levels of agreement for each item were calculated by averaging the priority they had assigned across the grading scenarios presented in the grading practices scale. A mean of 3.50 and above indicates principals' agreement to the scenario and a mean below 3.50 indicates principals' disagreement with the scenario. The overall descriptive statistics from Table 3 show principals' disagreement with most of the grading scenarios presented. For example, principals disagreed with the scenario of lowering students' grades as a result of their disruptive behaviors (Item 1; $M = 2.44$, $SD = 1.71$). Thus, principals did not generally consider students' behavior when grading. Again, principals agreed to award higher grades to students who showed active engagement in learning (i.e., completes homework, participates in classroom discussions, and answers questions) compared to those who were not active in learning (Item 7; $M = 4.04$, $SD = 1.46$). Moreover, principals' responses to the scenarios showed that their grading practices were not influenced by policy and external pressures. For example, principals disagreed to heeding to parents request to raise grades by 2% so that their child received an A in the course (Item 11; $M = 1.55$, $SD = .95$). They also did not endorse pressures from high authorities to inflate students' grades. They disagreed with the scenario of adhering to a consistent grade distribution of ensuring 30% of students received an A even if their test scores were not at an A level (Item 12; $M = 2.61$, $SD = 1.56$). Principals' grading was based on students' active engagement in learning.

Factor Analysis

Two sets of factor analyses were conducted, first in relation to the Conceptions of Assessment scale and then in relation to the grading scenarios. The factor analysis results for the principals' conceptions of assessment showed a three-factor model, with factors mirroring those found in previous administrations of the instrument. Means, standard deviations per scenario, and factor loadings are reported in Table 4. Factor 1 focused on summative forms of assessment. Principals perceived assessment as a way of measuring students' higher order thinking skills and determining what students have learned as an indicator of student and school success. This factor was termed, *summative assessment as accountability*. The second factor was termed, *assessment as irrelevant*. This factor related to assessment as unfair to students and interfering with teaching. Principals recognized that assessment could lead teachers to teach in a way against their beliefs. They believed assessment had imprecise processes and assessment results are just filed and ignored. The third factor was named, *assessment as improvement of teaching and learning*. This factor involved assessment as providing information that modifies ongoing teaching and acknowledges the significant role and impact of feedback towards student learning.

Table 3

Grading Practices Descriptive Statistics

Scenarios	Mean	SD	Skewness	Kurtosis
<i>To what extent do you agree or disagree with the Teachers' grading practice?</i>				
1. One of Mr. Smith's students has high academic ability as shown by his work. However, the student often has disruptive behaviors and interrupts the other students' learning. Mr. Smith decides to lower the student's grades.	2.44	1.71	.891	-.584
2. Due to the large class size, Ms. Thompson feels overwhelmed by the reality of marking a large number of assignments. Therefore, she regularly has students assessed through a multiple-choice test.	2.31	1.22	.770	.096
3. Students complain about Mr. Simpson's grades being lower than the grades they receive in other subject areas. Mr. Simpson explains to his students that his grades tend to be lower due to the difficulty of the content, and he maintains his grades.	3.77	1.61	-.271	-1.153
4. When Ms. Bird assigns her students' final course grades, she considers how a student is going through a rough time because of her parents' messy divorce. Ms. Bird increases this student's grades by a few points for a take-home assignment.	3.05	1.60	.191	-1.207
5. When it comes to recording students' final grades, Mr. Barton mostly relies on the results of tests. But sometimes his formative assessment information suggests that a student with lower scores has really understood the content better than the summative assessment	2.94	1.66	.306	-1.279
6. Ms. Johnson was talking to her colleagues about her classes and said how students' achievement was adequately represented by her assessments. When assigning grades, she would base her decisions on student achievement through various assessments rather than considering their effort.	2.14	.961	.888	1.408
7. Mr. Anderson values a positive classroom community. He awards higher grades to students who show active engagement in learning (i.e., completes homework, participates in classroom discussions, and answers questions).	4.04	1.46	-.508	-.722
8. Ms. Pearson has a student who is failing; however, Ms. Pearson believes the student will be devastated and give up if the student fails. Ms. Pearson decides to give the student the benefit of the doubt and awards a passing grade.	3.35	1.46	-.155	-1.103
9. Mr. Gagnon's students take five quizzes. When it comes time to assign final grades, Mr. Gagnon focuses on the most recent achievement and only counts the last 3 quizzes according to his department's policy.	3.06	1.42	.356	-.889
10. Ms. Cole determines her students' final grades based on a number of tests. She aligns her test content and question format with that of the provincial test. That way, she feels her students are better prepared for the provincial test.	3.52	1.42	-.021	-.958
11. A parent asked Mr. Roy to raise their daughter's grade by 2% so that she receives an A in the course. Mr. Roy takes the request into consideration and decides to increase the student's final grade.	1.55	.95	2.132	4.971
12. Ms. Landry's principal has made it clear to her that 30% of her students must receive an A even if their test scores are not at an A level. Ms. Landry complies.	2.61	1.56	.907	-.248

Table 3, continued

Scenarios	Mean	SD	Skewness	Kurtosis
<i>To what extent do you agree or disagree with the Teachers' grading practice?</i>				
13. Mr. Boucher is more lenient when assigning final grades to students in their final years of high school as they are applying for universities and colleges.	1.94	.99	1.146	1.109
14. One of Ms. Pelletier's students who wants to pursue a career in medicine has been working hard all year to achieve an 85% in Biology. After the final exam, the student's grade drops to 78%. Ms. Pelletier decides to raise the student's final grades to an 81%.	2.55	1.39	.792	-.268
15. One of Mr. Brown's students is failing English by 9%. Mr. Brown wants the student to advance with her friends to the next grade level. Mr. Brown decides to raise the final grade to a pass.	2.26	1.12	.889	.124
16. Ms. Wilson has a student in her class who is applying for a school award. However, the student's final grade is 1% lower than the requirements. Ms. Wilson knows how important this award is to the student and decides to raise the student's final grade.	2.28	1.17	.862	.027

Table 4

Factor Loadings of Conceptions of Assessment After Rotation

Statements	Factors		
	1	2	3
Q37 Assessment measures students' higher order thinking skills	.628		
Q40 Assessment results can be depended on.	.611		
Q22 Assessment results are trustworthy.	.604		
Q26 Assessment is an accurate indicator of a school's quality.	.529		
Q28 Assessment establishes what students have learned.	.518		
Q36 Assessment determines if students meet curriculum expectations.	.509		
Q39 Assessment allows different students to get different instruction.	.483		
Q23 Assessment forces teachers to teach in a way against their beliefs.		.652	
Q32 Assessment is unfair to students.		.636	
Q33 Assessment results are filed and ignored.		.584	
Q41 Assessment interferes with teaching.		.548	
Q43 Assessment is an imprecise process.		.464	
Q20 Assessment provides feedback to students about their performance.			.694
Q21 Assessment is integrated with teaching practice.			.611
Q29 Assessment provides feedback to students about their learning needs.			.563
Q30 Assessment information modifies ongoing teaching of students.			.475
Q34 Teachers should take into account the error in all assessment.			.420

Note. Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 6 iterations.

Table 5

Factor Loadings of Grading Practices After Rotation

Statements	Factors	
	Emotional Grading	Comprehensive Grading
15. One of Mr. Brown's students is failing English by 9%. Mr. Brown wants the student to advance with her friends to the next grade level. Mr. Brown decides to raise the final grade to a pass.	.751	
14. One of Ms. Pelletier's students who wants to pursue a career in medicine has been working hard all year to achieve an 85% in Biology. After the final exam, the student's grade drops to 78%. Ms. Pelletier decides to raise the student's final grades to an 81%.	.705	
13. Mr. Boucher is more lenient when assigning final grades to students in their final years of high school as they are applying for universities and colleges.	.696	
16. Ms. Wilson has a student in her class who is applying for a school award. However, the student's final grade is 1% lower than the requirements. Ms. Wilson knows how important this award is to the student and decides to raise the student's final grade.	.653	
11. A parent asked Mr. Roy to raise their daughter's grade by 2% so that she receives an A in the course. Mr. Roy takes the request into consideration and decides to increase the student's final grade	.589	
12. Ms. Landry's principal has made it clear to her that 30% of her students must receive an A even if their test scores are not at an A level. Ms. Landry complies.	.510	
4. When Ms. Bird assigns her students' final course grades, she considers how a student is going through a rough time because of her parents' messy divorce. Ms. Bird increases this student's grades by a few points for a take-home assignment.		.649
5. When it comes to recording students' final grades, Mr. Barton mostly relies on the results of tests. But sometimes his formative assessment information suggests that a student with lower scores has really understood the content better than the summative assessment		.560
9. Mr. Gagnon's students take five quizzes. When it comes time to assign final grades, Mr. Gagnon focuses on the most recent achievement and only counts the last 3 quizzes according to his department's policy.		.536
10 Ms. Cole determines her students' final grades based on a number of tests. She aligns her test content and question format with that of the provincial test. That way, she feels her students are better prepared for the provincial test.		.490
8. Ms. Pearson has a student who is failing; however, Ms. Pearson believes the student will be devastated and give up if the student fails. Ms. Pearson decides to give the student the benefit of the doubt and awards a passing grade.		.480

Note. Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 3 iterations.

The factor analysis results for the principals' grading practices revealed a two-factor model. Means, standard deviations per scenario, and factor loadings are reported in Table 5. The first

factor primarily centered on scenarios that resulted in inflation of students' grades based on principals' worries about the effects or consequences of the grades assigned to students. The first factor was named, *consequential grading*. These principals were concerned about students' failing, students being denied college or university admission, or not receiving school awards, and therefore are more likely to inflate their grades to avoid those consequences. These principals tended to respond to pressures from parents and higher authorities to inflate students' grades. Factor 2 focused on principals' multiple use of assessment indicators and consideration of assessment situations. This factor was termed, *comprehensive grading*. Principals agreed to making adjustments to grades based on their understanding of issues in students' homes and preferred using formative assessment evidence over summative assessments. They valued providing alternative opportunities for students to demonstrate their learning.

Correlations

To understand principal conceptions of assessment in relation to their grading practices, a Pearson correlation analysis was carried out. The correlation found significant positive relationship between comprehensive and consequential grading [$r = 0.436$, $n=141$, $p < 0.01$] (see Table 6). The result shows a significant positive correlation between consequential grading and those who perceive assessment as irrelevant [$r = 0.276$, $n=141$, $p < 0.01$]. This result implies that principals who perceive assessment as irrelevant are more likely to agree with consequential grading practices. If assessment is perceived as irrelevant with imprecise processes and is just filed and ignored, such grading practices may be more likely to inflate students' grades in order to avoid the negative effects or consequences of grades on students. Table 6 also shows a significant positive correlation between summative assessment as accountability and comprehensive grading [$r = 0.214$, $n = 141$, $p < 0.01$]. This result implies that principals who agree with accountability-oriented grading practices are more likely to agree with thorough grading practices. There is also a positive correlation between summative assessment as accountability and assessment as improvement of teaching and learning [$r = 0.515$, $n = 141$, $p < 0.01$]. Principals in this study conceived accountability and improvement function of assessment as inseparable.

Table 6

Pearson's Correlation Coefficients Correlation Between Conceptions of Assessment and Grading Practices

Variables	1	2	3	4	5
1. Consequential Grading	1				
2. Comprehensive Grading	.436**	1			
3. Summative assessment as accountability	.087	.241**	1		
4. Assessment as irrelevant	.276**	.006	-.260**	1	
5. Assessment as improvement of teaching and learning	-.049	-.064	.515**	-.250**	1

** . Correlation is significant at the 0.01 level (2-tailed).

Discussions

This study examined a limited sample of 141 Chinese school principals' conceptions of assessment broadly and grading practices specifically and the relationship between these two classroom practices. Considering the critical role that principals play in school effectiveness and student success globally, this study has important implications for both theory and practice. Our discussions first focus on principals' conceptions of assessment and then on their conceptions of grading practices.

Principals in this study believed that assessment is highly integrated with teaching and learning. Assessment is viewed as an effective way to improve teaching and learning as well as support accountability of students' learning and evaluation of school quality—a finding that has only been heightened by the COVID-19 pandemic. This finding is anchored in the correlation analysis of the three-factor model of principals' conceptions of assessment that showed a significant positive relationship ($r = 0.515$) between summative assessment as accountability and assessment as improvement of teaching and learning. Both assessment conceptions were viewed by principals as critically important to school and student performance, a finding supported globally based on the existing literature in China and Canada (Cheng, et. al., 2020; Davies et al., 2014; Renihan & Noonan, 2012). This finding is consistent with Brown et al.'s study (2009) in another Chinese context, but with teachers. Their study found that teachers highly associate ($r = 0.91$) summative assessment with formative assessment (i.e., assessment for improvement of teaching and learning). Within the Canadian context, Renihan and Noonan (2012) examined Saskatchewan rural school principals' assessment leadership roles and the impact of rural context on their work. They identified the tensions between the general demands of accountability on the one hand, and the changing leadership role demands presented by assessment reform and classroom assessment practices on the other. With the significant positive relationship between assessment for accountability and assessment for improvement, it appears that principals in this study believed holding students accountable for their learning through tests and examinations was a one way of assisting students to improve and become more responsible for their own learning. These findings are consistent with previous research conducted in the Chinese context that found Chinese teachers perceived students' grades on a test as a tool that could be used to motivate learning (Cheng et al., 2018). To this end, any kind of assessment that does not significantly contribute to accountability will not be viewed as contributing to improvement of teaching and learning (Brown & Gao, 2015). Our study's findings show a narrow yet intriguing way of understanding summative and formative functions of assessment, a value uniquely within the Chinese context yet echoed in assessment literature elsewhere in the world as stated earlier in the paper.

The conjoining of formative, summative, and accountability functions of assessment is not surprising in the Chinese system of education as China's educational assessment is highly dominated by high-stakes tests and examinations, especially at the end of middle school and senior secondary school (Zhong Kao 中考 and Gao Kao 高考 respectively; Cheng, 2008; Brown & Gao, 2015; He et al., 2011; Li, 2009; Liu & Qi, 2005; Niu, 2007). As mentioned at the beginning of this paper, student performance in public examinations is used to evaluate the quality of teachers and schools, and the overall capacity of students, but also teachers' incomes are linked to student examination results (Harris et al., 2009; Zhang & Ng, 2011). Since quality of teachers and schools are largely determined by students' performance in the public examinations, principals tend to prioritize maximizing students' performance on the externally administered

tests. In lieu of this, it will be substantially difficult for principals to endorse or support implementation of assessment that avoids using examinations. Assessment that does not significantly contribute to accountability would not be viewed as contributing to improvement of teaching and learning in this context (see also Brown & Gao, 2015). This assertion is supported by a cross-national study conducted by Birenbaum et al. (2015). The study found that demand for accountability has mounted pressure on school systems worldwide, consequently becoming one of the largest impediments to achieving improvements associated with formative assessment. Although the side effects of large-scale assessments as accountability are well understood in education (see Baidoo-Anu & Ennu Baidoo, 2022; Emler et al., 2019), understanding such a function as perceived by the principals in this study as also related to improvement of teaching and learning can be viewed, in some contexts, as contradictory (Zhao, 2014).

The high-stakes examination acceptance and accountability pressures have shaped principals' view of summative and formative assessment as inseparable but has also influenced some principals, in this study, to perceive assessment as irrelevant. Brown and Gao (2015) indicated that the reasons for conceiving assessment as irrelevant differed between Chinese and non-Chinese teachers. Non-Chinese teachers gave reasons in relation to the impact of assessment on the individual students and its inherent flaws, Chinese teachers, on the other hand, gave explanations in relation to high-stakes public examinations, and the linking of teachers' incomes to the performance of students in those examinations (Harris et al., 2009; Zhang & Ng, 2011). Since principals' conceptions of assessment in this study are similar to those of the teachers in the Chinese context (Brown et al., 2009, Li & Hui, 2007), high-stakes examinations and accountability demands might be a plausible explanation for principals' conception of assessment as irrelevant in this educational context.

In this era of assessment-driven teaching and learning globally, having principals who hold negative views of assessment could be problematic, and is an area that needs further research. Principals who conceive assessment as irrelevant may not be able to decipher and engage quality alignment of teaching, learning, and assessments and appropriately interpret and use the results of assessment to improve teaching and learning (Arter et al., 1993). It is also important to explore further the contextually-appropriate use and reporting of student achievement information within the Chinese education system in a way that supports and moves learning forward. Consideration for an accountability framework that appropriately repositions large-scale assessment to support, but not control school improvement is worth exploring (Volante, 2007). The synergy (or lack thereof) that often exists between large-scale testing and teachers' assessment practices is often the result of contradictory messages given to school leaders and classroom practitioners (Birenbaum et. al, 2015). Therefore, a balance in principals' conceptions of assessment system will support effective implementation of assessment in teaching and learning and obtain quality information about the progress and success of their school (Jakicic, 2009).

Moreover, in term of grading practices, the study revealed principals' degree of agreement to grading and how it relates with their assessment conceptions. Factor analyses of principals' agreement of grading practices in this study revealed a two-factor model—consequential grading and comprehensive grading. Principals are concerned about the effect or consequences of (consequential) grade use. They are worried about the high-stakes nature of grade use such as, students being denied of college or university admission at home and abroad and receiving school awards or not. Similarly, Tierney (2015) also found that teachers in other parts of world altered students' grades based on: (a) caring and compassion, (b) a consideration of student future

opportunities, and (c) an attempt to give life lessons to students.

Correlational analyses of principals' conceptions of assessment and grading practices reveals that principals who are more concerned about the use of students' grade may be more likely to conceive assessment as irrelevant. As a result, they may be more likely to agree with inflating students' grades to avoid those consequences. Moreover, because they conceive assessment as irrelevant amidst accountability pressures, they may even be more likely to easily succumb to pressures from parents and higher authorities to inflate students' grades to avoid consequences of grade use. This finding resonates with previous work by Cheng et al. (2018). Although Chinese teachers experienced pressure from parents and students to inflate grades, they were mostly pressured by school administrators to inflate students' grades. This is because school administrators in China are mostly expected to maintain a particular percentage of As in their schools (especially top high schools) which attracts first class students into their schools.

Again, principals who agreed with a comprehensive approach to grading consider multiple factors such as understanding of issues at students' homes, and students' active engagement in learning (participating in classroom discussions and answering questions) as the best way to grade students. Correlational analyses of principals' conceptions of assessment and grading practices further revealed that principals' conception of assessment as summative and accountability may likely influence them to be more comprehensive in their grading. This is not surprising in a society and culture like China where students' performance in examinations is linked to their overall worth as individuals and used for making high-stake decisions. Although China has enacted grading policies that serve to promote comprehensive quality education (Ministry of Education of the People's Republic of China, 2002; 2013), accountability pressures continue to make principals more thorough with their approach to grading. This finding supports the works of (Cheng et al., 2004; Cheng et al., 2008; Sun & Cheng, 2013) who investigated teachers' assessment and grading practices in Canada, Hong Kong, and China. These studies show that teachers' grading preferences are influenced by their values about assessment.

Conclusions

This current study addresses a timely research gap related to principals' conceptions of assessment and grading practices within the Chinese educational contexts. Although it is well understood that principals have the responsibility of being instruction leaders in this context (Li et al., 2012; Lunenburg, 2010; Marzano & Waters, 2010), principals (and educators), in this context and globally, continue to wrestle with what this role means in relation to assessment leadership and responsibilities. The continued emphasis of assessment in schools is part of a global paradigm shift in which student assessment is viewed as essential in enhancing instruction and student learning. This study thus has practical implications for the professional development of principals in the areas of assessment, seen as a broad instructional practice, and grading, as a specific summative and evaluative instructional practice.

Due to the global COVID-19 pandemic, our data were collected from a restricted sample in a huge country, China, thus limits the generalizability of the study's findings, which is a limitation of the study. Therefore, future research could increase the sample size to allow for broader generalization of the results and build in specific questions to explore the impact of global pandemic on the alignment of teaching, learning, and assessment in a drastically different manner. Furthermore, our data were limited to the use of a survey design to a finite number of scales and items, which did not provide insights into why principals viewed assessment and

grading practices in certain ways in their contexts. Further qualitative research is recommended to investigate further the drivers of principals' conceptions of assessment and grading practices.

Notwithstanding, this study has provided some unique compelling evidence on how principals navigate assessment and grading in the examination-oriented culture of China. Understanding the socio-cultural setting where teaching, learning, assessment is situated is essential if we are to fully understand the embedded values and beliefs that determine the instructional practices in context across educational systems. As a significant amount of literature focuses on teachers' conceptions of assessment and grading, examining principals' assessment conceptions in relation to their grading practices is critical to support consistent assessment and grading practices in schools. Principals' leadership in student assessment is a catalyst for productive and positive assessment in schools; ultimately enhancing school effectiveness and student success.

Acknowledgement

This work was supported by an Insight Grant from the Social Sciences and Humanities Research Council (SSHRC) of Canada. We also want to thank Dr. Wei Yan at the Faculty of Education, Queen's University and Professor Rong Yan at ILEAD, Xian Jiaotong-Liverpool University for their work and support in the data collection of the project.

References

- Alberta Teachers' Association. (2004). *Leadership quality standard*.
<https://www.alberta.ca/assets/documents/ed-leadership-quality-standard-english.pdf>
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218–223.
<https://doi.org/10.3200/TCHS.78.5.218-223>
- Arter, J. A., Stiggins, R. J., Duke, D., & Sagor, R. (1993). Promoting assessment literacy among principals. *NASSP Bulletin*, 77(556), 1–7. <https://doi.org/10.1177/019263659307755602>
- Baidoo-Anu, D., & Ennu Baidoo, I. (2022). Performance-based accountability: Exploring Ghanaian teachers perception of the influence of large-scale testing on teaching and learning. *Education Inquiry*, <https://doi.org/10.1080/20004508.2022.2110673>
- Ball, C. M. (2016). *How elementary school principals manage accountability expectations*. [Doctorial Dissertation, The University of Western Ontario]. The University of Western Ontario Electronic Thesis and Dissertation Repository.4328. <https://ir.lib.uwo.ca/etd/4328>
- Bandalos, D. L., & Finney, S. J. (2018). Factor analysis: Exploratory and confirmatory. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 98–122). Routledge. <https://doi.org/10.4324/9781315755649-8>
- Barnes, N., Fives, H., & Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107–116. <https://doi.org/10.1016/j.tate.2017.02.017>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Birenbaum, M., Deluca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volente, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140.
<https://doi.org/10.1177%2F1478210314566733>
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and Learning* (pp. 81–100). Sage.

- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. <https://doi.org/10.1177/003172171009200119>
- Bonner, S. M., & Chen, P. P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, 14(2), 57–77. <https://doi.org/10.1080/10627190903039411>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brookhart, S. M. (2013). Grading. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 257–272). Sage.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166–170. <https://doi.org/10.2466/pro.99.1.166-170>
- Brown, G. T. L. & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes, *Cogent Education*, 2(1), 1–19. <https://doi.org/10.1080/2331186X.2014.993836>
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multi-Disciplinary Evaluation*, 6(12), 68–91. https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/236
- Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50(5-6), 307–320. <https://doi.org/10.1016/j.ijer.2011.10.003>
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice*, 16(3), 347–363. <https://doi.org/10.1080/09695940903319737>
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27(1), 210–220. <https://doi.org/10.1016/j.tate.2010.08.003>
- Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology Education*, 26, 319–337. <https://doi.org/10.1007/s10212-010-0052-3>
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15(1), 75–89. https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37286
- Carless, D. (2010). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings* (1st ed.). Routledge. <https://doi.org/10.4324/9780203128213>
- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20(4), 369–383. <https://doi.org/10.1191/0265532203lt2640a>
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25(1), 15–38. <https://doi.org/10.1177/0265532207083743>
- Cheng, L., & Curtis, A. (Eds.). (2009). *English language assessment and the Chinese learner* (1st ed.). Routledge. <https://doi.org/10.4324/9780203873045>
- Cheng, L., DeLuca, C., Braund, H., & Yan, W. (2018). Grading policies in Canada and China: A comparative study. *Comparative and International Education / Éducation Comparée et Internationale*, 47(1), Article 5. <https://doi.org/10.5206/cie-eci.v47i1.9324>
- Cheng, L., DeLuca, C., Braund, H., Yan, W., & Rasooli, A. (2020). Teachers' grading decisions and practices across cultures: Exploring the value, consistency, and construction of grades across

- Canadian and Chinese secondary schools. *Studies in Educational Evaluation*, 67. <https://doi.org/10.1016/j.stueduc.2020.100928>
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing*, 21(3), 360–389. <https://doi.org/10.1191/0265532204lt288oa>
- Cheng, L., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9–32. <https://doi.org/10.1080/02602930601122555>
- Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12(2), 213–233. <https://doi.org/10.1080/15434303.2015.1010726>
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107. <https://doi.org/10.1080/15434300701348409>
- China Civilization Centre. (2007). *China: Five thousand years of history and civilization*. City University of Hong Kong Press.
- Daniels, L. M., & Poth, C. A. (2017). Relationships between pre-service teachers' conceptions of assessment, approaches to instruction, and assessment: An achievement goal theory perspective. *Educational Psychology*, 37(7), 835–853. <https://doi.org/10.1080/01443410.2017.1293800>
- Davies, A., Busick, K., Herbst, S., & Sherman, A. (2014). System leaders using assessment for learning as both the change and the change process: Developing theory from practice. *Curriculum Journal*, 25(4), 567–592. <https://doi.org/10.1080/09585176.2014.964276>
- DeLuca, C. (2011). Interpretive validity theory: Mapping a methodology for the validation of educational assessments. *Educational Research*, 53(3), 303–320. <https://doi.org/10.1080/00131881.2011.598659>
- DeLuca, C., Cheng, L., & Volente, L. (2019). *Grading across Canada; Policies, practices, and perils*. <https://www.edcan.ca/articles/grading-across-canada/>
- Deluca, C., Donaldson, G., Hayward, L., Tan, K., & Wyatt-Smith, C. (2021). *Imperatives for a better assessment future during and post COVID*. IEAN: International Educational Assessment Network. <https://iean.network/gallery/iean-assessment-imperatives-covid-may2021.pdf>
- Doucet, A., Netolicky, D., Timmers, K., & Tuscano, F. J. (2020). *Thinking about pedagogy in an unfolding pandemic: An independent report on approaches to distance learning during the COVID-19 school closures*. <https://learningportal.iiep.unesco.org/en/library/thinking-about-pedagogy-in-an-unfolding-pandemic-an-independent-report-on-approaches-to>
- Emler T. E., Zhao Y, Deng J, Yin D, Wang Y. (2019). Side effects of large-scale assessments in education. *ECNU Review of Education*, 2(3), 279–296. <https://doi.org/10.1177/2096531119878964>
- Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, 107(1), 186–213. <https://doi.org/10.1111/j.1467-9620.2005.00462.x>
- Field, A. (2013) *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock “N” roll*, (4th ed.). Los Angeles, London: Sage.
- Gebril, A., & Brown, G. T. L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment Education Principle. Policy Practice*, 21(1), 16–33. <https://doi.org/10.1080/0969594X.2013.831030>
- Hargreaves, E. (2013). Inquiring into children's experiences of teacher feedback: Reconceptualising Assessment for Learning. *Oxford Review of Education*, 39(2), 78–94. <https://doi.org/10.1080/03054985.2013.787922>
- Harris, J., Zhao, Y., & Caldwell, B. J. (2009). Global characteristics of school transformation in China. *Asia Pacific Journal of Education*, 29(4), 413–426. <https://doi.org/10.1080/02188790903308860>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- He, Y., Levin, B. B., & Li, Y. (2011). Comparing the content and sources of the pedagogical beliefs of

- Chinese and American pre-service teachers. *Journal of Education for Teaching*, 37(2), 155–171. <https://doi.org/10.1080/02607476.2011.558270>
- Herppich, S., Praetorius, A., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>
- Hutcheson, G. & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage. <https://doi.org/10.4135/9780857028075>
- Jakicic, C. (2009). A principal's guide to assessment. In T. R. Guskey (ed.), *The principal as assessment leader* (pp. 53–72). Solution Tree Press.
- Kennedy, K. J. (2007, 28–30 May). Barriers to innovative school practice: a socio-cultural framework for understanding assessment practices in Asia. In K. J. Kennedy, P. Fok, F. W. M. Yu, & J. K. S. Chan, *Student Assessment and its Social and Cultural Contexts: How Teachers Respond to Assessment Reforms*[Symposium]. Redesigning Pedagogy—Culture, Understanding and Practice Conference, Singapore. https://www.eduhk.hk/fpece_project/QEF/Work/Papers%20prepared%20for%20the%20Symposium_edited.pdf
- Kennedy, K. J. (2016). Exploring the influence of culture on assessment: The case of teachers' conceptions of assessment in Confucian-heritage cultures. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 404–419). Routledge. <https://doi.org/10.4324/9781315749136>
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment Education: Principles, Policy & Practice*, 16(3), 263–268. <https://doi.org/10.1080/09695940903319646>
- Kohn, A. (1999). Grading is degrading. *Education Digest*, 65 (1), 59–64. <https://www.alfiekohn.org/article/degrading-de-grading/>
- Lentz, D. L. N. (1997). *The assessment, grading and reporting practices of selected elementary school teachers and principals* (Accession No. 9825006). [Doctoral dissertation, University of Texas, Austin]. University of Texas, Austin ProQuest Dissertations Publishing.
- Li, B. R. (2012). Using formative assessment to promote learner autonomy in college English teaching. *Journal of Hebei United University (Social Science Edition)*, 12(5), 189–191.
- Li, J. (2009). Learning to self-perfect: Chinese beliefs about learning. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 35–69). Hong Kong: Springer/The University of Hong Kong, Comparative education Research Centre.
- Li, W. S., & Hui, S. K. F. (2007). Conceptions of assessment of mainland China college lecturers: A technical paper analyzing the Chinese version of CoA-III. *The Asia-Pacific Education Researcher*, 16, 185–198. <https://ejournals.ph/article.php?id=3856>
- Li, X. L., Li, W. H., & Lu, N. G. (2012). Cujin jiaoshi fazhan de xiaozhang lingdaoli yanjiu [Research on the principals' leadership for promoting teacher development: The case study of two schools' principals]. *Jiaoshi Fazhan Yanjiu (Educational Development)*, 4, 70–74.
- Liu, P., & Qi, C. (2005). Reform in the curriculum of basic education in the People's Republic of China: Pedagogy, application, and learners. *International Journal of Educational Reform*, 14(1), 35–44. <https://doi.org/10.1177/105678790501400103>
- Liu, X., & Yan, W. (2016). Grading in East Asia: An examination of government policies. *Graduate Student Symposium Selected Papers*, 10, 73–96. <http://hdl.handle.net/1974/14036>
- Lo, K. P. (2004). Adaptation and development of schooling in changing societies. In R. Koo, C. K. Kam, & M. S. Yung (Eds.) *Education & curriculum reform adaptation and development in the Pearl River Delta Region*. The Association for Childhood Educational International.

- Lunenburg, F. C. (2010). The principal as instructional leader. *National Forum of Educational and Supervision Journal*, 27(4), 1–7. <https://pdf4pro.com/view/the-principal-as-instructional-leader-national-54fac.html>
- Ma, J., & Cheng, L. (2016). Chinese student' perceptions of the value of test preparation courses for the TOEFL iBT: Merit, worth and significance. *TESL Canada Journal*, 33(1), 58–79. <https://doi.org/10.18806/tesl.v33i1.1227>
- Marzano, R. J. (2000). *Transforming classroom grading*. Association for Supervision and Curriculum Development.
- Marzano, R. J., & Waters, T. (2010). *District leadership that works: Striking the right balance*. Solution Tree.
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results* [Google Books version]. Retrieved on April 23, 2020, from <https://books.google.com/>
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203–213. <https://doi.org/10.1080/00220670209596593>
- Ministry of Education of the People's Republic of China. (2002). 教育部关于积极推进中小学评价与考试制度改革的通知 [Notice on promoting the reform of the assessment and examination systems of elementary and secondary schools]. http://www.moe.gov.cn/srcsite/A26/s7054/200212/t20021227_166074.html
- Ministry of Education of the People's Republic of China. (2012). “中小学校长培训规定” [Training requirements for primary and secondary school principals]. www.gov.cn/fwxx/bw/jyb/content_2267061.htm
- Ministry of Education of the People's Republic of China. (2013). 教育部关于推进中小学教育质量综合评价改革的意见 [Guidelines on promoting the comprehensive assessment of educational quality in elementary and secondary schools]. http://www.moe.gov.cn/srcsite/A26/s7054/201306/t20130608_153185.html
- Morrison, K. (2006). Paradox lost: Toward a robust test of the Chinese learner. *Educational Journal*, 34(1), 1–30.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25. <https://doi.org/10.1111/j.1745-3992.2003.tb00140.x>
- National Bureau of Statistics of China (2014). *China Statistical Yearbook*. China Statistics Press.
- Nitko, A. J. (2001). *Educational assessment of students*. 3rd ed. Merrill/Prentice Hall.
- Niu, W. (2007). Western influences on Chinese educational testing. *Comparative Education*, 43(1), 71–91. <https://doi.org/10.1080/03050060601162412>
- O'Connor, K. (2007). *A repair kit for grading: 15 fixes for broken grades*. Pearson.
- OECD. (2013). *Synergies for Better Learning: An international perspective on evaluation and assessment. Pointers for policy development*. <http://www.oecd.org/site/eduistp13/Synergies%20for%20Better%20Learning%20-%20Policy%20Pointers.pdf>
- OECD. (2016). *Education in China: A snapshot*. OECD Publishing. <https://www.oecd.org/china/Education-in-China-a-snapshot.pdf>
- Ontario Ministry of Education. (2010). Growing Success: Assessment, evaluation, and reporting in Ontario schools. <http://www.edu.gov.on.ca/eng/policyfunding/growSuccess.pdf>
- Paine, L. W. (1990). The teacher as virtuoso: A Chinese model for teaching. *Teachers College Record*, 92(1), 49–81. <https://doi.org/10.1177/016146819009200105>
- Parker, M. (2006). *How principals' beliefs about classroom assessment influence their leadership practices: An exploration* (Accession No. MR36234). [Masters Thesis, University of Lethbridge, Lethbridge]. University of Lethbridge, Lethbridge ProQuest Dissertations Publishing. <https://hdl.handle.net/10133/542>

- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Putnam, R. T., & Borko, H. (1997). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29(1), 4–15. <https://doi.org/10.3102%2F0013189X029001004>
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380. <https://doi.org/10.1016/j.tate.2010.03.008>
- Renihan, P., & Noonan, B. (2012). Principals as assessment leaders in rural schools. *The Rural Educator*, 33(3), 1–8. <https://doi.org/10.35608/ruraled.v33i3.407>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>
- Simon, M., Chitpin, S., & Yahya, R. (2010). Pre-service teachers' thinking about student assessment issues. *The International Journal of Education*, 2(2), 1–22. <https://doi.org/10.5296/ije.v2i2.490>
- Stiggins, R. J., & Chappuis, J. (2006). What a difference a word makes: Assessment FOR learning rather than assessment OF learning helps students succeed. *Journal of Staff Development*, 27(1), 10–14. <https://www.jaymctighe.com/wp-content/uploads/2011/04/Assessment-for-Learning.pdf>
- Stiggins, R. J., & Duke, R. L. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285–291. <https://doi.org/10.1177%2F003172170809000410>
- Sun, Y., & Cheng, L. (2013). Teachers' grading practices: Meanings and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. <https://doi.org/10.1080/0969594X.2013.768207>
- Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, 8, 5–30. <https://doi.org/10.18296/am.0002>
- Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum*, 75(3), 210–227. <https://doi.org/10.1080/00131725.2011.577669>
- Torrance, H. (2001). Assessment for learning: Developing formative assessment in the classroom, *Education 3-13*, 29(3), 26–32. <https://doi.org/10.1080/03004270185200331>
- Trompenaars, F. & Hampden-Turner, C. (1998). *Riding the waves of culture: Understanding cultural diversity in business*. Nicholas Brealey Publishing. https://ocan.yasar.edu.tr/wp-content/uploads/2013/09/Riding-the-waves_Part-1.pdf
- Vandeyar, S., & Killen, R. (2007). Educators' conceptions and practice of classroom assessment in post-apartheid South Africa. *South African Journal of Education*, 27(1), 101–115. <https://www.ajol.info/index.php/saje/article/view/25101/20770>
- Volante, L. (2007). Education quality and accountability in Ontario: Past, present, and future. *Canadian Journal of Educational Administration and Policy*, 58, 2–21. <https://journalhosting.ucalgary.ca/index.php/cjeap/article/view/42739>
- Watkins, M. W. (2000). *Monte Carlo PCA for parallel analysis (computer software)*. State College, PA: Ed & Psych Associate.
- Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, 34(3), 14–27. <https://www.jstor.org/stable/41064580>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Zhang, X. F., & Ng, H. M. (2011). A case study of teacher appraisal in Shanghai, China: In relation to teacher professional development. *Asia Pacific Education Review*, 12, 569–580. <https://doi.org/10.1007/s12564-011-9159-8>
- Zhao, Y. (2014). *Who's afraid of the big bad dragon? Why China has the best (and worst) education system in the world*. Wiley.

Liyang Cheng (PhD) is Professor in Language Education and Assessment at the Faculty of Education, Queen's University. Her seminal research on washback illustrates the global impact of large-scale testing on instruction, the relationships between assessment and instruction, and the academic and professional acculturation of international and new immigrant students, workers, and professionals in Canada—an underrepresented population in the research, but one whose lives are greatly impacted by testing and assessment. Liyang focuses her research on using assessment to support student and teacher learning involving, in particular, linguistically and culturally diverse people around the world.

David Baidoo-Anu (him/his) is currently a PhD Candidate at the Faculty of Education, Queen's University in Kingston, Ontario Canada. He earned his Master of Philosophy (MPhil) in educational measurement and evaluation from the University of Cape Coast-Ghana. David's current research focuses on promoting equitable assessment and learning experiences for underrepresented groups of students through culturally responsive classroom assessment. Particularly, he draws on culturally situated understandings of assessment and empirical evidence showing the positive impacts of quality classroom assessment practices on raising students' achievement to advocate for systemic assessment reforms. The aim of his doctoral research is to support quality and consistent assessment practices in schools across cultural context and ultimately enhance the assessment experiences of all students regardless of their background.

Christopher DeLuca (PhD, Queen's University) is a Professor and an Associate Dean at the School of Graduate Studies and Postdoctoral Affairs and Professor in Educational Assessment at the Faculty of Education, Queen's University. Dr. DeLuca leads the Classroom Assessment Research Team and is Director of the Queen's Assessment and Evaluation Group. Dr. DeLuca's research examines the complex intersection of assessment, curriculum, and pedagogy as operating within the current context of school accountability and standards-based education. His work largely focuses on supporting teachers in negotiating these critical areas of practice to enhance student learning experiences. In particular, Dr. DeLuca's research centers on how pre-service and in-service teachers learn to engage the complexities of assessing student learning in relation to the evolving accountability culture in today's classrooms.