Exploring assessment across cultures: Teachers’ approaches to assessment in the U.S., China, and Canada

Christopher DeLuca¹, Nathan Rickey¹* and Andrew Coombs¹

Abstract: Classroom assessment dynamics are shaped by individual and local understandings of assessment (assessment micro-cultures), as well as common assessment beliefs and practices that stem from system-wide features, such as large-scale testing (assessment macro-cultures). Teachers’ approaches to assessment reveal how they navigate assessment micro- and macro-cultures to support student learning and achievement. Despite increasing migration of students between the U.S., China, and Canada, little research has examined the different approaches to assessment students encounter when they move between these contexts. Thus, the specific supports they need to adapt to their new assessment cultures and have equitable access to learning have remained unclear. This exploratory research compared teachers’ approaches to assessment in the U.S., China, and Canada. Latent class analysis identified five types of assessors across these contexts: teacher-centric assessors, hesitant assessors, moderately student-
centric assessors, highly student-centric assessors, and eager assessors. Associations between assessor type and country were identified, revealing different patterns in how teachers approach assessment in each education context.

**Subjects:** Teaching & Learning - Education; Educational Research; Education Studies; Multicultural Education; Assessment & Testing; Education Policy; Assessment

**Keywords:** Approaches to Assessment; Assessment Macro-Culture; Assessment Micro-Culture; Latent Class Analysis; Assessment Policy; Large-Scale Assessment; Assessment Literacy

1. Introduction

Globalization has led to an upsurge in migration of students between education contexts. A dominant trend over the past two decades has been students from China studying in the United States (U.S.) and Canada, representing 33.7% (Institute of International Education, 2019) and 34.1% (Cheng et al., 2018; Statistics Canada, 2016; Zhihua & Kumari, 2014) of all international students, respectively. Students who move between these jurisdictions can face obstacles when adapting to new education contexts, such as adapting to a student-centered teaching style. Research has suggested that unfamiliar assessment practices and conceptions form a significant barrier to their full inclusion in new education systems (Li, 2004; Liu et al., 2016; Zhang & Zhou, 2010) as teachers understand and implement assessment differently across education systems and sociocultural contexts (Brown et al., 2019; Kennedy, 2016). Given the wealth of empirical research that highlights assessment’s key role in student learning (e.g., Birenbaum et al., 2015; Herppich et al., 2018; McMillan, 2013), it is unsurprising that students’ success in school is contingent upon their acculturation to new ways of thinking about assessment when they move between education contexts. However, cross-cultural research comparing the different assessment cultures in the U.S., China, and Canada is scarce, meaning that specific supports necessary to support students who move between these contexts are unclear.

The culture of assessment within an education system can be understood on two levels. First, patterns in how teachers understand assessment fairness, purposes, and grading protocols have been attributed to systemwide features, such as jurisdictional assessment policies, large-scale testing frameworks, and post-secondary admissions, as well as sociocultural traditions embedded within system boundaries (Brown & Harris, 2009; Brown et al., 2009; Cheng et al., 2018; DeLuca et al., 2020; Liu et al., 2016; Xu & Brown, 2016). The construct of assessment macro-cultures, which includes educational and sociocultural traditions that shape common assessment beliefs and practices across a geographical space (Allal, 2016), encapsulates the influence of these processes.

However, at the classroom level, teachers are required to engage in a dynamic process of navigating and discerning local school, classroom, and cultural knowledge with students and colleagues in order to use assessment in the service of students’ learning (Willis et al., 2013). Highly context-dependent, this ongoing dynamic amounts to unique assessment micro-cultures defined by the assessment beliefs, practices, and tools in each classroom (Allal, 2016). Research has shown that assessment macro-culture features of an education system directly or distally shape some common assessment conceptions and practices across teachers within the education context (Brown & Harris, 2009; Fulmer et al., 2015; Xu & Brown, 2016). While they are influenced by assessment macro-cultures, assessment micro-cultures are primarily shaped by teachers’ and students’ understandings of the relationship between assessment, teaching, and learning (Allal, 2016; DeLuca et al., 2020; Tierney, 2006; Yung, 2002).

Teachers’ approaches to assessment—their theoretical and philosophical orientations towards assessment purposes, processes, fairness and theory (DeLuca et al., 2016b)—offer a comprehensive understanding of how teachers navigate both assessment macro- and micro-
cultures to respond to contemporary assessment situations. In other words, teachers’ approaches to assessment tap into how they respond to both systemwide features and more localized understandings of assessment. Patterns in how teachers approach assessment across an educational context thereby offer insight into (a) the ways in which students are likely to experience assessment within an education context, and (b) specific barriers students face as they move from one educational context to another. Perhaps most importantly, examining teachers’ assessment approaches can build on sociocultural understandings of classroom assessment by excavating how assessment is understood and enacted across three different sociocultural contexts.

Despite their importance in students’ learning, little is known about the different approaches to assessment students encounter when they move between China, the U.S., and Canada. Empirical studies have suggested differences related to teachers’ assessment practice by jurisdiction; however, research has only focused on specific elements of assessment (e.g., fairness), and there is a lack of cross-cultural evidence in the three contexts of interest. Consequently, the specific supports that can be provided to these students in adapting to new assessment cultures remain unclear. Given the high number of students who move between the U.S., Canada, and China, this gap in research leaves many unprepared to meet their full potential in school. To support students who move between these contexts, research examining patterns in how teachers navigate and negotiate assessment macro- and micro-cultures in different national contexts to implement classroom assessment is critically needed. Accordingly, the purpose of this exploratory study was to provide initial insights into teachers’ approaches to assessment in the U.S., China, and Canada. Guiding this research were the following questions:

1. Are there distinct patterns in teachers’ approaches to classroom assessment in the United States, China, and Canada?
2. Do the prevalence of these patterns differ by country?

1.1. Assessment macro-cultures: the U.S., China, and Canada
Examining and comparing assessment approaches in the U.S., China, and Canada begins by understanding systemwide features that characterize each assessment macro-culture. Assessment macro-culture features (e.g., teacher accountability systems, education policy) influence how teachers implement assessment across an education context (Brown & Harris, 2009; Fulmer et al., 2015; Xu & Brown, 2016). While quantifying or specifying causal links between macro- and micro-cultural features is beyond the scope of this article, having a broad understanding of key distinctions and synergies between the U.S., Chinese, and Canadian assessment macro-cultures provides the context for interpreting the assessment approaches that emerge in each context. Although assessment macro-cultures in part shape assessment micro-cultures (i.e., teachers’ assessment beliefs and practices; Brown & Harris, 2009; Xu & Brown, 2016), we discuss these as discrete constructs in order to frame findings from relevant research. The intention of the following section is not to suggest that one homogenous education system exists within each country context; in fact, this paper will later explore several micro-cultural features that shape how education is practiced and explains some differences that exist within systems. Instead, here we provide a broad overview of each macro-culture in order to contextualize our findings and explore certain countrywide features that may partly shape teachers’ approaches to assessment.

1.2. United States
The education system in the U.S. is decentralized and primarily the responsibility of each of the 50 states. Most educational policies, including curriculum and assessment frameworks, are decided at state and local levels (U.S. Department of Education, n. d.). However, the federal Department of Education’s Every Student Succeeds Act (ESSA; 2015) provides an overarching framework for state and local policymakers. The ESSA emphasizes equitable access to education and standards-based accountability measures.
The ESSA extends a tradition of using statewide testing systems to evaluate and improve the performance of schools. Large-scale testing in the U.S. is generally focused on school and teacher accountability rather than improving student learning (Nichols & Harris, 2016). Linking students’ statewide test results to sanctions and pay incentives (Committee on Incentives and Test-Based Accountability in Public Education, 2011; Lashway, 2001; Stobart & Eggen, 2012) has resulted in widespread teaching to the test, narrowed curricula emphasizing tested subjects, a focus on measuring learning, and reliance on efficient testing modalities (e.g., multiple-choice tests; Birenbaum et al., 2015; Hamilton et al., 2012; McMillan, 2013; Stiggins, 2002).

Researchers have advocated for greater emphasis on classroom assessment practices that support, rather than merely test, students’ learning. Shepard (2000, 2005) has called for a shift in the U.S. assessment macro-culture from a testing culture, focused on using assessment to summate student learning, to a culture wherein assessment is primarily used to support learning. Policymakers have given increasing attention to formative assessment (e.g., National Research Council, 2001; The Gordon Commission, 2013), with state policy more frequently citing the definition of The Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards (FAST SCASS; Birenbaum et al., 2015; FAST SCASS, 2006; Gordon et al., 2014). FAST’s work shifted focus away from a psychometric perspective on formative assessment validity (Shepard, 2009) and instead emphasized that formative assessment is an ongoing process of adjusting teaching and learning (Perie et al., 2007). To bridge the gap between day-to-day formative assessment and infrequent state or district wide summative assessments, districts and/or schools implement interim assessments to diagnose learning needs, inform instruction, provide timely feedback to students, and predict outcomes on summative assessments (O’Keefe & Lewis, 2019; Perie et al., 2007). Despite this three-tiered assessment framework, the U.S.’s ongoing tradition of standards-based accountability rooted in psychometric measurement principles has eroded trust in teachers using professional discretion to make pedagogical adjustments during instruction. Consequently, teachers have tended to use formative assessment as a measurement tool (e.g., diagnostic quizzes; Erickson, 2007).

1.3. China

China’s Ministry of Education governs education using countrywide curriculum and grading policies to direct provincial and municipal ministries and schools (Wang, 2012; Xu, 2004). China’s highly centralized education system is shaped by its long tradition of public examinations (Brown & Gao, 2015; Cheng et al., 2018; Xu, 2004). Critical gatekeeping exams have significant social consequences for students and teachers (e.g., quality of post-secondary education and accountability measures, respectively). Pressure to prepare students for these large-scale exams efficiently has resulted in large class sizes, lecture style pedagogy, teaching to the test, and excessive exam practice (Chen & Brown, 2013; Y. Gu, 2006; Liu et al., 2016; Yin & Buck, 2015). Researchers have raised concerns that students in China may primarily engage in surface learning and rote-memorization tactics, with little opportunity to develop critical thinking skills or confident understandings of complex content (Gan et al., 2017; Zhou & Zhang, 2014).

In recognizing the potential negative impact high-stakes testing can have on student learning, the Ministry of Education has called for a classroom assessment system that, instead of promoting transmission and rote memorization of knowledge for success on large-scale exams, encourages teachers to employ assessment practices that drive authentic learning (Chen & Brown, 2013; Jin & Cortazzi, 2006; OECD, 2011). Although large-scale exams are still primarily selection and gatekeeping mechanisms, a recent initiative sees teacher-developed classroom assessments integrated into the public examination system (Gan et al., 2017). More importantly, curricula reforms emphasize formative assessment, or Assessment for Learning (AfL) over summative assessment (i.e., assessment to summate or certify learning). Policymakers in China articulate AfL as classroom assessment practices designed to provide feedback to teachers and students to guide teaching and learning. Further, AfL should enhance students’ abilities to regulate their learning autonomously and thus includes student self- and peer assessment (Chen & Brown, 2013; Gan et al.,
Policymakers increasingly call for student-centered assessments, wide varieties of assessment tasks to monitor and improve learning, integration of assessment tasks with daily learning expectations, and recognition of the value of teachers' observations (Zheng, 2017). However, despite policymakers' efforts to encourage effective use of AFL in China (The Ministry of Education of P. R. China, 2012; Zheng, 2017), research has suggested that there is limited use of AFL in classrooms (Zhao et al., 2017).

1.4. Canada

Education in Canada is decentralized and is primarily the responsibility of the provinces and territories (Klinger et al., 2008; Volante & Jaafar, 2008). Unlike China and the U.S., Canada does not have a national policy framework outlining a countrywide curriculum. The provincial systems have their own policies overseeing educational standards, curriculum, classroom practices, and assessment. These policies share a similar social constructivist understanding of learning that values collaboration and inquiry approaches (Peterson & McClay, 2010). Further, provincial/territorial policy frameworks are intended to be interpreted by local school boards or districts (Volante & Jaafar, 2008) using educators' professional judgment (e.g., Ontario Ministry of Education, 2010).

Assessment policies across Canada's 13 systems similarly delineate the purposes of classroom assessment. They emphasize implementation of AFL to provide ongoing feedback to students and support their learning (Birenbaum et al., 2015; Cheng et al., 2018) and highlight tasks that actively engage students in the assessment process (e.g., peer and self-assessment; Alberta Education, 2006; British Columbia Ministry of Education, 2004). For example, The Ontario Ministry of Education (2010) specifies a sub purpose of AFL called Assessment as Learning (AoL) which refers to assessment tasks that engage students' metacognition. Thus, the policies highlight a student-centered conceptualization of AFL. Policies also direct teachers to design and use summative assessment (also Assessment of Learning, AoL) to grade students' performance against established standards (e.g., Ontario Ministry of Education, 2010). Grades in Canada only represent students' academic achievement; teachers report non-achievement factors (e.g., behavior, initiative, organization) separately (DeLuca et al., 2017; Tierney et al., 2011). Grading is primarily the teacher's responsibility as it is based on classroom AoL.

While each province has its own large-scale testing system, education ministries use the results of provincial tests primarily to monitor and improve system effectiveness (Klinger et al., 2008; Klinger & Rogers, 2011). Provincial test results are not used to hold teachers accountable or select students for higher education programs. They affect between 10–50% of students' grades in some provinces and grades (e.g., Alberta; Alberta Education, 2019; DeLuca et al., 2017). Despite their comparatively limited impact on teachers and students in Canada, large-scale testing frameworks have been found influence teaching and assessment practices (e.g., teaching to the test; Volante, 2010; Volante & Beckett, 2011).

Broadly summarizing these assessment macro-cultures reveals key distinctions among them. The U.S. assessment macro-culture is largely characterized by school and teacher accountability via statewide testing. In China, however, the assessment macro-culture is focused on student accountability and competitive large-scale exams. In the Canadian assessment macro-culture, large-scale testing systems differ greatly across the provinces. However, large-scale testing systems generally focus on system accountability while the emphasis for teachers is placed on the integration of AFL practices into instruction.

1.5. Assessment micro-cultures

The assessment macro-cultures described above in part shape the ongoing, dynamic social and cultural interactions that distinguish assessment micro-cultures found in individual classrooms. However, students' experiences of assessment and learning are not defined by policies and large-scale testing frameworks, but by how these policies and frameworks are interpreted and enacted by teachers. The assessment beliefs, practices, and tools students encounter and engage with are
shaped by how their teachers navigate macro-cultures as well as the teacher’s and students’ socially and culturally constructed understandings of assessment. While insights into individual classrooms are rare, empirical findings permit some insight into how macro-cultural elements translate to trends across assessment micro-cultures in the U.S., China, and Canada.

First, we review literature that offers insights into patterns across assessment micro-cultures in each context separately; then, we review cross-cultural research conducted in any combination of the three contexts of interest.

1.6. The United States

In a recent study by Johnson et al. (2019), examinations of teachers’ formative assessment practices across three educational districts in the U.S. revealed that teachers tended to implement questioning and learning tasks effectively but needed support in sharing learning criteria, providing individual feedback in lessons, and fostering collaboration. Teachers’ reliance on questioning and implementing learning tasks, coupled with the noted lack of peer and self-assessment, supported previous findings that teachers in the U.S. tend to employ formative assessment as a teacher-controlled measurement instrument. Additionally, research has examined teachers’ ethical dilemmas in grading and found that approximately two-thirds of reported grading dilemmas were related to score pollution (i.e., invalid measurement of students’ mastery), particularly involving standardized testing and special populations (Pope et al., 2009). Collectively, this research suggests that measurement is a central consideration for teachers in the U.S. However, research has also shown that U.S. teachers across subject areas used assessment that focused on students’ individual development rather than comparing students or teachers’ grade distributions (McMillan et al., 2003).

Examining teachers’ conceptions of the purposes of assessment in the U.S., Barnes et al. (2017) found that teachers aligned with one of three conceptions: assessment (a) as valid for school accountability, (b) improves teaching and learning, and (c) is irrelevant. Items that had been assigned to assessment makes students accountable and assessment improves education conceptions in other assessment macro-cultures instead united in the assessment as valid for accountability conception in the U.S. The authors posited that this difference reflected the effects of the U.S.’s macro-culture of assessment accountability on teachers’ interpretations of items. These findings suggest that teachers in the U.S., like those in China, value both summative and formative assessment yet distrust the assessment’s capacity to measure or inform learning.

1.7. China

Empirical research has revealed that summative assessment has remained a focus for teachers in China (Cai et al., 2017). Other findings have indicated that many teachers in China conceptualized assessment as merely an accountability mechanism, irrelevant to learning processes and dependable as a basis for instructional decision-making (Brown & Gao, 2015; Zhao et al., 2016). Researchers have reported a strong association between the conceptualization of assessment as a formative process and as an accountability mechanism, suggesting that teachers who believed the primary purpose of assessment is to drive students’ learning forward also believed teachers should be held accountable for students’ results on summative examinations (Brown et al., 2011). The persistent emphasis of summative assessment in China may be due to large-scale testing pressures (Brown & Gao, 2015; Gu, 2014).

Moreover, in Confucian heritage societies, achievement on public examinations is seen as a reflection of the quality of one’s character. Therefore, cultural norms in China dictate that preparing students well for large-scale examinations is tantamount to making them better people, providing a moral impetus for focusing on summative assessment (Brown et al., 2011; Kennedy, 2016). Confucian heritage cultural norms also shape how teachers in China implement formative assessment. As both teacher authority and knowledge transmission are valued, teachers tend to implement formative assessment tasks that generate feedback for the teacher rather than ones
that harness student autonomy (Poole, 2016). Examining teachers’ approaches to assessment in China, Coombs et al. (2021) found that teachers in China prioritized approaches to measurement theory (i.e., consistent, contextual, and balanced approaches).

1.8. Canada

Empirical research investigating how teachers implement classroom assessment in Canada offers insight into teachers’ assessment tools, feedback methods, purposes, and conceptualizations of assessment fairness. Examining English teachers’ classroom assessment practice through the assessment tools they use (e.g., essays, portfolio entries, tests), Hunter et al. (2006) found that participants tended to use assessment tools to facilitate one-to-one interactions to support students’ learning. Peterson and McClay (2010), investigating assessment and feedback practices, similarly found that grade 4–8 English teachers were highly concerned with how they used assessment to facilitate communication with students to support their learning. Teachers reported using motivating comments, oral feedback, and teacher-supported peer feedback. However, portfolios were not widely used to demonstrate student growth or foster students’ self-assessment skills. Further, teachers voiced a concern for making English assessments objective by using well-designed measurement tools (e.g., rubrics). Taken together, these findings suggest that the participants used assessments that give students an active role in the assessment process to some extent but prioritized dependable measurement of learning to facilitate teacher-driven interventions (e.g., discussions with students).

Other scholars have noted growing but still limited integration of approaches that engage student agency in assessments. Examining how AoL, Afl, and AaL were enacted in Ontario schools, Volante (2010) argued that teachers place excessive emphasis on AoL and noted barriers that limit teachers’ implementation of AaL (e.g., peer and self-assessment). Later research examining Ontario teachers’ formative assessment practices noted a shift in focus from grading to teaching wherein teachers emphasized providing feedback without grades (Volante & Beckett, 2011). While findings suggested that while there were still barriers to AaL implementation, teachers recognized the importance of making students active participants in assessment processes.

More recent research examining how pre-service teachers conceptualize the purpose of classroom assessment supports that teachers in Canada are increasingly adopting the stance that assessment improves learning. Research investigating pre-service teachers’ conceptions of assessment purposes found that pre-service teachers in the province of Alberta more highly endorsed positive conceptions of assessment, such as “assessment improves learning” and “assessment improves teaching” compared to negative conceptions (e.g., assessment is ignored, assessment is bad; Daniels et al., 2014). Despite high support for positive conceptions, this study also found that participants expressed higher endorsement for one negative conception—assessment is inaccurate—compared to other countries. While the authors speculate that high endorsement of assessment is inaccurate could be due to various reasons related to the participants being pre-service teachers (e.g., they were learning about measurement and validity issues when surveyed), this conception could also be specific to sample from Alberta, the province with the greatest use of large-scale assessment, as other research has suggested that teachers in Canada trust classroom assessments over large-scale assessments for providing accurate cognitive diagnostic information (Leighton et al., 2010). Other research studying Alberta pre-service teachers identified two distinct approaches to assessment as understood through an achievement goal theory perspective. Daniels and Poth (2017) noted that pre-service teachers in their sample more highly endorsed mastery approaches to assessment (i.e., assessment approaches that focus on learning rather than demonstrating competence) compared to performance approaches (i.e., assessment approaches that signal that demonstrating competency and achieving high grades is more important than learning). This finding would suggest that the participants emphasize Assessment for Learning over preparing for graded assessments. Coombs et al. (2020) studied pre-service teachers’ approaches to assessment in Canada and grouped participants using a latent class analysis. They found that the largest class endorsed contemporary assessment approaches, such as
Assessment as Learning and tailoring assessments to meet the individual needs of students. However, the extent to which findings from these studies generalize to teachers throughout Canada is limited because pre-service teachers likely hold idealistic conceptions of assessment (Daniels & Poth, 2017).

Research on grading has shown that teachers in Canada used achievement as the primary means of determining grades; however, in borderline cases, teachers considered students’ strong effort and good behavior (i.e., achievement-related factors) as justification for giving a higher grade (Duncan & Noonan, 2007; Hunter et al., 2006). Tierney (2014) found that teachers’ conceptions of fair assessment emphasized providing opportunities for all to demonstrate their learning; addressing individual students’ learning needs rather than using standard assessments; and transparent communication of learning expectations. Taken together, empirical research suggests that teachers in Canada (a) value communication in assessment as an aspect of fairness and effective teaching, (b) increasingly conceptualize assessment as improving teaching and learning, and (c) may attempt to implement AoL but are limited in their capacity to do so.

Evaluating the research examining assessment practices in the separate contexts of the U.S., China, and Canada reveals few points of comparison because researchers have focused on different dimensions of assessment (e.g., assessment purpose, fairness) and used different measures to examine assessment practices. Collectively, research conducted in each context could suggest that accountability pressures in all three contexts drive a focus on measurement reliability and validity despite teachers’ efforts, particularly in Canada, to adopt more student-centered approaches. However, such assertions must be approached with caution, as making valid comparisons between contexts requires cross-cultural research.

1.9. Comparative perspectives
The limited comparative research examining how assessment is practiced by classroom teachers in the U.S., China, and Canada has focused primarily on ethics and fairness in assessment and grading. Liu et al. (2016) found significant differences in perceptions of ethical classroom assessment practices between preservice teachers in China and the U.S., highlighting fundamental differences in how educators think about assessment ethics shaped by embedded sociocultural traditions. Findings suggested that educators in the U.S. and China both recognized supporting students’ achievement in large-scale testing as paramount to ethical assessment practice; however, the classroom assessment practices they endorsed in preparing students for large-scale tests differed. Educators in the U.S. emphasized standardized approaches to measurement of learning, while educators in China endorsed considering non-achievement factors (e.g., giftedness) to shape ethical assessment practice.

Comparative research has also explored different grading practices across the U.S., China, and Canada. Assessment policy documents in China direct teachers to grade holistically: grades in China represent a combination of academic achievement and non-achievement factors (e.g., effort, initiative, behavior). In contrast, assessment policies across Canada mandate that achievement and non-achievement factors are reported separately so that grades represent only academic achievement (Cheng et al., 2018, 2017). Liu et al. (2016) results suggested that policy to grade holistically does inform practice in China, where almost all participants believe considering students’ effort and growth in grading was ethical practice. Seventy percent of the U.S. participants also perceived considering non-achievement factors when assigning grades as ethical, suggesting that holistic grading may be common in the U.S. as well. Tata (2005) examined the extent to which students valued interpersonal justice (i.e., respectful treatment) and voice (i.e., opportunities to discuss and appeal grading decisions) in grading in the U.S. and China. Teachers in both countries valued interpersonal justice and voice in grading; however, U.S. participants perceived the absence of voice as less fair than China participants, while participants in China were more concerned with the absence of interpersonal justice. The authors attributed these perspectives on fairness to differing sociocultural traditions. Cross-
cultural findings therefore suggest that sociocultural norms in the U.S., China, and Canada result in differing perceptions of grading fairness and sources of evidence that inform grading. Given the importance of grades for selection purposes, different grading procedures and understandings of fairness can present significant challenges and consequences for students who move between these educational contexts (Cheng et al., 2018; Liu et al., 2016).

Ultimately, education policies and available classroom assessment research offer some insight into how educators navigate assessment macro- and micro-cultures in the U.S., China, and Canada, particularly with regards to assessment purposes and grading practices. However, empirical research examining classroom assessment practices within each context has mainly focused on teachers’ conceptions of the purposes of assessment (e.g., Barnes et al., 2017; Brown & Gao, 2015), assessment ethics (e.g., Pope et al., 2009), or fairness (e.g., Tierney, 2014) in isolation. While it has begun to highlight underlying juxtapositions, research comparing classroom assessment across these contexts is sparse and has focused on grading policies (e.g., Cheng et al., 2018), teachers’ perceptions of assessment fairness (e.g., Tata, 2005), and preservice teachers’ views on ethical assessment (e.g., Liu et al., 2016). Accordingly, two gaps emerge from our evaluation of the relevant literature. First, there is a need for cross-cultural research that investigates and draws comparisons between teacher samples in the U.S., China, and Canada using a consistent measure of assessment practice. Second, as prior research has tended to focus on one aspect of assessment (e.g., fairness, purpose, process) in isolation, research is needed that takes a holistic, multidimensional view of assessment practice in the identified contexts. Such research would reveal how endorsement of various dimensions of classroom assessment (e.g., purpose, fairness, process) interrelate, illuminating students’ experiences of assessment and therefore learning. Our research addresses both gaps by examining teachers’ approaches to assessment in the U.S., China, and Canada.

1.10. Approaches to classroom assessment

While researchers have operationalized several constructs to gain insight into teachers’ assessment practices, approaches to assessment conceptualizes assessment practices as a multidimensional construct, providing a holistic, comprehensive view of how teachers understand and practice assessment. Deluca and colleagues (Coombs et al., 2018; DeLuca et al., 2016b, 2018) operationalize approaches to assessment as teachers’ philosophical and theoretical orientations towards 12 assessment dimensions that shape how they practice classroom assessment within their sociocultural and policy contexts. Defined by their responses to authentic classroom assessment scenarios on the Approaches to Classroom Assessment Inventory (ACAI), teachers’ approaches to assessment illuminate their endorsement of 12 approaches related to assessment purposes, processes, fairness, and measurement theory collectively, whereas previous research has investigated aspects of assessment in the U.S., China, and Canada in isolation. Further, teachers can adhere to multiple or even incongruent approaches. See Table 1 for an explanation of each approaches to assessment dimension.

As a measure of how teachers prioritize multiple dimensions of assessment in contemporary classroom assessment scenarios, approaches to assessment tap into how teachers negotiate assessment macro- and micro-cultures in their assessment practice. Specifically, teachers’ responses to the contemporary classroom assessment scenarios are informed by consideration of their specific classroom, school, and broader educational contexts. Teachers’ approaches to assessment are thus situated at the intersection of macro- and micro-cultural influences. Examining teachers’ approaches to assessment across the U.S., China, and Canada has the potential to reveal patterns in how students experience assessment differently in each context. Identifying patterns in teachers’ approaches to assessment is thus necessary to supporting students who move between these macro-cultures in adapting to new assessment micro-cultures and overcoming specific barriers they are likely to encounter.
2. Methods
A survey method was used to analyze classroom teachers’ approaches to assessment in the U.S., China, and Canada. In total, responses from 710 teachers were analyzed. All responses were anonymous. Participants indicated their informed consent on an online letter of information and consent form and then completed the online survey. Ethical approval was obtained by the General Research Ethics Board at Queen's University.
2.1. Sample
Participants were contacted and recruited through LISTSERVs maintained by national, provincial, regional, and state organizations and through social media advertising. In total, 710 classroom teachers were included in the analysis, relatively equally distributed across national contexts: the U.S. (n = 227); China (n = 250); and Canada (n = 233). Participants were from a range of states and provinces within their respective contexts. In the U.S., participants were from four states: New Jersey (55.2%), Michigan (33.0%), South Carolina (6.1%), and Pennsylvania (5.7%). In China, participants were from Shandong province (37.5%), Zhejiang province (25.5%), Heilongjiang province (20.1%), and Jilin province (16.9%). Participants in Canada were from the Eastern Provinces (Newfoundland and Labrador, Nova Scotia, New Brunswick, and Prince Edward Island; 56.8%), the Central Provinces (Quebec and Ontario; 26.6%), and the Western Provinces (Saskatchewan, Alberta, and British Columbia; 15.9%). Country samples were similar in terms of proportions of genders and ages, and participants were from a range of teaching divisions. Participants in the U.S. taught in a mix of public and private schools, while participants in China taught in federal public schools and teachers in Canada generally taught in public schools. See Table 2 for further demographics.

2.2. Instrument
Participants in the U.S., China, and Canada completed the Approaches to Classroom Assessment Inventory (ACAI) which contains demographic items and four ACAI scenarios. The ACAI was designed to enable teachers to understand and map their approaches to classroom assessment (DeLuca et al., 2016a). The 12 approaches dimensions were deduced based on analysis of 15 contemporary assessment standards across five countries (DeLuca et al., 2016b; for the full instrument, see, 2016a). See Table 1 for definitions of each dimension.

The focus of the ACAI is four scenario-based items that were developed using an expert panel methodology (DeLuca et al., 2018). Ten educational assessment experts from across North America examined each of the scenarios and actions to ensure they aligned with the underlying approaches to assessment construct. The alignments of the scenarios and actions were verified by 10 practitioner experts (five elementary and five secondary teachers). To date, the ACAI has been

| Demographic variable | U.S. (n = 227) | China (n = 250) | Canada (n = 233) |
|----------------------|---------------|----------------|-----------------|
| Certification*       |               |                |                 |
| Elementary           | 67.8          | 17.2           | 23.1            |
| Middle School        | 33.9          | 51.6           | 69.0            |
| High School          | 31.3          | 39.2           | 53.8            |
| Gender               |               |                |                 |
| Male                 | 23.5          | 31.3           | 20.9            |
| Female               | 74.3          | 67.6           | 77.0            |
| Other/no response    | 2.2           | 0.0            | 2.1             |
| Age                  |               |                |                 |
| <25                  | 6.2           | 5.2            | 18.1            |
| 26–30                | 14.2          | 16.5           | 11.5            |
| 31–40                | 33.4          | 43.8           | 24.3            |
| 41–50                | 26.5          | 30.1           | 27.0            |
| >51                  | 19.7          | 4.4            | 19.1            |

Note: A small number of teachers were certified to teach multiple levels of school.
used in various contexts, including Germany, Canada (DeLuca et al., 2020), the U.S. (Barnes et al., 2020), and China (Coombs et al., 2021). It shows promise for understanding how assessment is approached in a variety of policy and socio-cultural contexts. Further, the ACAI has proven useful in comparing approaches across education contexts (e.g., DeLuca et al., 2020).

Each of the four ACAI scenarios represents a contemporary classroom assessment dilemma in which multiple defensible actions can be taken. Participants are asked to interpret the scenarios in relation to their current teaching context (i.e., grade, school, community). After reading each scenario, participants rated the likelihood of taking 12 defensible actions using 6-point Likert scales (1 = Highly Unlikely—6 = Highly Likely) with an option to select “Don’t Know”. Participants could similarly endorse any number of actions. Each of the 12 actions aligned with one of the 12 approaches to assessment dimensions. By rating their likelihood of taking each action, participants expressed their level of support for each of the 12 assessment dimensions in each of the four scenarios. Endorsement of a particular approach was determined by averaging each participant’s support for a specific dimension across all four scenarios (DeLuca et al., 2020). For example, Scenario 1 presented the following situation: You give your class a paper-pencil summative unit test with accommodations and modifications for identified learners. Also, 16 of the 24 students fail. For Scenario 1, participants rated their likelihood of taking actions such as have students generate a plan to relearn the material. As this action relates to the dimension of Assessment as Learning (AaL), participants who highly rated this action support an AaL approach in this scenario.

2.3. Data analyses
Descriptive statistics (e.g., mean, standard deviation) were calculated for all items. Latent class analysis (LCA) was performed using the 12 approaches to assessment within the ACAI as indicators (models for classes 3–5 can be seen in Table 4). LCA is a technique used to analyze relationships in categorical data in order to identify underlying class/group membership among participants based upon their joint conditional probabilities of endorsement for a collection of items (McCutcheon, 1987). Unlike traditional cluster analysis techniques which group participants based on arbitrarily chosen criteria, LCA employs maximum likelihood methods to estimate group membership probabilities and provides goodness-of-fit indices to enable comparisons between latent class models (Kim et al., 2018). LCA has proven to be an effective method of classifying participants based on their characteristics (Kim et al., 2018; Zhang et al., 2017). Building on previous work in the area of teachers’ approaches to and practice of classroom assessment (Coombs et al., 2020; Veldhuis & Van Den Heuvel-panhuizen, 2014), a LCA approach was chosen.
Table 3. Descriptive statistics for all participants (n = 710)

| Assessment dimension | All participants | United States (n = 227) | China (n = 250) | Canada (n = 233) | Internal consistency (Cronbach’s alpha) |
|----------------------|------------------|--------------------------|----------------|-----------------|----------------------------------------|
| Assessment of learning | 3.79 (1.19)      | 4.28 (.72)               | 3.43 (.89)     | 2.79 (1.31)     | .535                                   |
| Assessment for learning | 4.56 (.95)      | 4.59 (.73)               | 4.28 (.98)     | 4.50 (1.35)     | .682                                   |
| Assessment as learning | 4.07 (1.06)      | 4.00 (.86)               | 4.21 (.99)     | 4.27 (1.40)     | .679                                   |
| Design               | 4.43 (1.96)      | 4.48 (.69)               | 4.06 (1.03)    | 4.30 (1.40)     | .629                                   |
| Use/scoring          | 3.55 (1.02)      | 3.68 (.80)               | 3.50 (.91)     | 3.30 (1.35)     | .487                                   |
| Communication        | 4.35 (1.99)      | 4.26 (.83)               | 4.35 (.93)     | 4.57 (1.28)     | .674                                   |
| Standard             | 3.54 (1.05)      | 3.90 (.74)               | 3.69 (.90)     | 2.76 (1.16)     | .544                                   |
| Equitable            | 3.84 (1.19)      | 4.25 (.73)               | 4.04 (.92)     | 2.98 (1.46)     | .617                                   |
| Differentiated       | 3.97 (1.06)      | 4.00 (.81)               | 4.00 (.99)     | 3.89 (1.46)     | .625                                   |
| Consistent           | 4.02 (1.01)      | 4.11 (.66)               | 3.82 (.88)     | 3.81 (1.50)     | .464                                   |
| Contextual           | 3.96 (1.00)      | 3.90 (.66)               | 3.96 (1.02)    | 4.11 (1.50)     | .579                                   |
| Balanced             | 4.16 (1.02)      | 4.24 (.73)               | 4.05 (.96)     | 4.00 (1.45)     | .608                                   |

Note: 1 = not at all likely, 6 = highly likely.

To allow for examination of how participants, rather than items, grouped together. Further, LCA could reveal patterns in how the 12 approaches interrelate and manifest in students’ classroom assessment experiences, whereas comparing mean differences between individual approaches (e.g., assessment fairness) would reveal very little about teachers’ overall approaches. Endorsement of 12 approaches to assessment were used to group participants into classes that shared similar levels of support for a particular combination of approaches to assessment. LCA was performed using 1000 random starting values to generate the best loglikelihood that could be replicated. Similar research on teachers’ approaches to assessment (Coombs et al., 2020; Timmons & Pelletier, 2016; Veldhuis & van den Heuvel-Panhuizen, 2014) have identified 3–4 latent classes. Consequently, this study tested models that ranged from 2 to 6.

The best-fitting latent class model was determined by examining absolute (Vuong–Lo–Mendell–Rubin likelihood ratio test [VLMLR-LRT] and the Lo–Mendell–Rubin adjusted likelihood ratio test [LMR-LRT]) and relative (BIC, SSA-BIC) fit indices (Muthén & Muthén, 2000; Nylund, Asparouhov, & Muthén, 2007). VLMR-LRT and LMR-LRT provide information regarding absolute model fit between k and k-1 models with a significant p-value indicating better fit for k class model means compared to the k-1 class model (Lo et al., 2001). BIC and SSA-BIC provide information on relative model fit, with lower values indicating better model fit. As these are BIC and SSA-BIC indicate relative model fit, no hard cutoff (i.e., statistical significance) is employed when evaluating these statistics. The entropy value ranges from 0 to 1, with higher values indicating smaller classification error for the model (Celeux & Soromenho, 1996). Entropy values close to or exceeding .800 are generally accepted as noteworthy (Roesch et al., 2010). These statistics, in addition to class size and interpretability considerations, informed the selection of the model (Lanza et al., 2007).

Classes were named according to conditional probabilities (Figure 1; Clark & Muthén, 2009; Nylund, Bellmore, Nishina, & Graham, 2007). The presence of latent classes was used to indicate if there were distinct patterns teachers’ approaches to assessment with the conditional probabilities within each class used to explain the characteristics of each class.
Using a 3 × 4 contingency table (cross tables), Pearson’s chi-square tests of independence (chi-square test) were used to determine if there were any significant associations between class membership and country. A chi-square test using z-test of column proportions with Bonferroni adjustments to significance level (α = 0.05) was employed to identify significant differences between class memberships. LCA was conducted using Mplus version 7.4 (Muthén & Muthén, 2015) while all other data analyses were completed using the Statistical Program for the Social Sciences v. 22 (SPSS).

3. Results

Table 3 provides overall descriptive statistics for teachers’ approaches to assessment. Across all participants, the lowest endorsed approaches were use/scoring of assessment (M = 3.55, SD = 1.03) and a standard approach to assessment fairness (M = 3.54, SD = 1.05) while the highest endorsed approaches were Assessment for Learning (M = 4.56, SD = .95) and design of assessments (M = 4.43, SD = .96). These trends were generally observed in each country sample. However, teachers in China and Canada also endorsed communication approaches relatively highly (M = 4.35, SD = .93 and M = 4.57, SD = 1.28, respectively). Further, AoL approaches were among the least endorsed approaches in China (M = 3.43, SD = .89) and Canada (M = 2.79, SD = 1.31).

The internal consistencies for the 12 subscales ranged between .464 and .682. While the internal consistencies of these subscales are relatively low, they are in line with similar previous studies examining assessment practices of in-service teachers (Xu & Brown, 2017). In their study of in-service teachers in China, Xu and Brown (2017) posited that their low internal consistency was the result of few items representing each construct dimension and because in-service teachers have sufficient experience to distinguish between constructs. Like in Xu and Brown’s study, our subscales are comprised of responses to relatively few items (i.e., four) and we studied in-service teachers in China (in addition to the U.S. and Canada), meaning that our observed internal consistencies were likely subject to the same influences as Xu and Brown’s study. Further, the ACAI scenario items were developed via an extensive analysis of contemporary assessment standards across five countries and subjected to assessment expert and practitioner panel reviews (DeLuca et al., 2016b). The ACAI has been used in research published in a variety of contexts (e.g., Coombs et al., 2020; DeLuca et al., 2020; Barnes et al., 2020).

3.1. Latent class analysis

A five-class model was a robust statistical fit for the data with class sizes that allowed for a statistical examination of class membership. The VLMR-LRT and LMRA-LRT noted a statistically significant better fit for the five-class model compared to the four-class model. The entropy value for a four-class model was moderately high (0.787), suggesting there was precision in assigning individuals to their appropriate classes. Further, differences entropy for all classes examined was similar, ranging from .787 to .847. Additionally, the rate of decline in the relative fit statistics (BIC, SSA-BIC) slowed at the
Table 5. Relationship between Latent Class Membership and Country of Residence as shown by country prevalence (%) within each class

| Country            | United States | China | Canada |
|--------------------|---------------|-------|--------|
| Teacher-centric assessors | 71.4 a        | 9.5 b | 19.0 b |
| Hesitant assessors  | 2.9 a         | 44.1 b| 52.9 b |
| Moderately student-centric assessors | 7.1 a     | 30.3 b| 62.6 c |
| Highly student-centric assessors | 35.6 a    | 36.8 a| 27.6 b |
| Eager assessors     | 52.2 a        | 42.9 b| 4.9 c  |

Note. “a”, “b”, and “c” indicate statistically significant difference at p < .05 confidence level in rows (i.e., country differences). Given the relatively small overall sample size, some of the classes include small numbers of teachers. We chose to maintain the 5-class model to enable comparisons across countries. However, the small sample size is a limitation of this study and further research is needed to interrogate the extent to which these classes are represented across each country.

five-class model, suggesting that this was the best-fitting model for the data (see Table 4). The five distinct latent classes were named: (a) teacher-centric assessors (5.9%); (b) hesitant assessors (4.8%); (c) moderately student-centric assessors (29.7%); (d) highly student-centric assessors (33.7%); and (e) eager assessors (25.9%). Figure 1 presents conditional probabilities of endorsing each dimension for each of the five classes.

3.2. Teacher-centric assessors

Teacher-centric assessors are those teachers who regularly engage in teacher-led and focused assessment. They primarily value assessment as a process to provide feedback and information to enhance their own pedagogical practice. While they endorse both AFL and AoL highly, teacher-centric assessors’ high endorsement of design, consistent, and standard approaches suggest they are primarily concerned with designing and implementing assessments that reliably measure learning. Coupled with their low endorsement of AoL and differentiated approaches, these assessors are likely to use formative assessment primarily as a strategy to inform future instruction rather than support students in using assessment to inform their own learning processes. In particular, AoL, use and scoring of assessments, and a differentiated approach to assessment fairness were not well endorsed by these teachers.

3.3. Hesitant assessors

Hesitant assessors were unlikely to endorse any of the approaches to assessment as all conditional probabilities were lower than 50%. Relatively higher probabilities of support for standard and consistent approaches, coupled with no likely support for equitable approaches, indicate a relative concern for reliable, consistent measurement of learning across groups. However, low endorsement of design approaches suggests low concern for creating assessments that reliably measure learning. This contradiction may indicate that hesitant assessors value reliability but do not feel qualified to design them and thus invest little effort in design. Another contradiction emerged in hesitant assessors’ approaches to assessment purpose. No endorsement of AFL and relatively high support for AoL indicate a skepticism towards assessment’s capacity to inform learning and a view that assessment is for summating learning. However, AoL received relatively high support, suggesting that hesitant assessors believe students should take an active role in assessment processes, particularly when it drives metacognitive learning forward. Thus, while they feel assessment may not be a trustworthy guide for teaching, hesitant assessors see a dual purpose of assessment: to enhance students’ metacognition and certify learning. Contradictory approaches to assessment reliability and purpose may hinder their capacity to measure and enhance student learning.
In any case, low endorsement for all assessment approaches suggests that hesitant assessors mistrust classroom assessment as both a mechanism for improving learning and collecting evidence capable of generating credible inferences about student learning. Hesitant assessors likely spend minimal time on assessment and instead focus on other aspects of teaching and learning (e.g., lesson planning) which they see as independent from assessment. These assessors are likely to align with an “assessment as irrelevant” conception of classroom assessment as identified in research conducted in China (Brown & Gao, 2015; Brown et al., 2011) and the U.S. (Barnes et al., 2017), or possibly with intuitive (Gipps et al., 1995) or head-noting (Hill, 2000) perspectives on assessment described in other countries.

3.4. Moderately student-centric assessors
Moderately student-centric assessors reported moderate endorsement of all approaches to assessment as conditional probabilities clustered around 50%. Relatively higher support for AFL, AoL, communication, and balanced approaches, combined with lower support for AoL and standard fairness, reflect the general approaches of highly student-centric assessors (described below). However, the most likely to be endorsed approach, AFL, only received moderate support: 60.2%. Further, more than half of the approaches were unlikely to be endorsed. These findings may suggest that moderately student-centric assessors adhere to contemporary assessment standards and policies (e.g., Klinger et al., 2015), but that they do not feel strongly about classroom assessment’s role in teaching and learning. Their approaches may be informed by the assessment standards they were taught in contemporary teacher education programs; however, they reflect a more realistic view of their capacity to integrate these standards into their practice given time constraints and pressures of full-time teaching (e.g., administrative duties, behavior management, preparing students for large-scale tests).

3.5. Highly student-centric assessors
Highly student-centric assessors strongly endorsed assessment approaches that support students’ growth and development, and that accounted for individual students’ learning needs and contexts. Specifically, these assessors highly endorsed the following dimensions: AFL, AoL, design and communication of assessments, equitable and differentiated approaches to fairness, and balancing consideration of assessment context with the need for assessment consistency across contexts. In the classroom, the assessment practices of these teachers are likely to reflect contemporary standards and assessment policies in Canada and United States (e.g., The Classroom Assessment Standards for PreK–12 Teachers; Klinger et al., 2015; Ontario Ministry of Education, 2010). Conditional probabilities for the remaining approaches to assessment clustered lower, suggesting that contemporary assessors are less likely to endorse them. These approaches included AoL, use and scoring of assessments, a standard and differentiated approach to fairness, and valuing a consistent approach to classroom assessment.

3.6. Eager assessors
Eager assessors highly endorsed all 12 approaches to assessment across all five scenarios, despite supporting seemingly contradictory approaches within the same scenario. An example of endorsing contradictory approaches would be a teacher who prioritized adapting rubrics and scoring guides to reflect identified students’ learning needs and using the same scoring rubric for all students without applying criteria differently based on individual student needs. In the classroom, these teachers are likely to invest a great deal of time and energy into a wide range of assessment practices; however, these practices may have counter effects on learners and the learning climate. Further, it is possible that eager assessors simply highly endorsed all actions to provide what they perceived as socially desirable responses.

4. Statistical differences between classes
The chi-square tests showed significant associations between class membership and country ($X^2 [8, N = 710] = 217.213, p < 0.0001$) (see Table 5). The only group which showed no significant differences between any country was highly student-centric assessors.
4.1. The United States
Teachers in the U.S. tended to group into three classes: teacher-centric, eager, and highly student-centric assessors. With more than twice the expected proportional class membership, teachers in the U.S. were significantly more likely to be teacher-centric assessors than teachers in China or Canada. A significant association was also found between teachers in the U.S. and eager assessors, with teachers in the U.S. more likely to be eager assessors than teachers in China and Canada. Last, while no significant differences were found between countries, more U.S. teachers than expected were grouped with highly student-centric assessors.

Because teaching in the U.S. was not associated with membership in hesitant or moderately student-centric assessor groups, these findings first suggest that teachers in the U.S. highly value assessment. Students in the U.S. are likely to encounter teacher-centric, eager, and highly student-centric approaches, meaning that assessment will always be a significant part of their learning experience. In navigating assessment macro- and micro-cultures in the U.S., teachers either highly control assessment processes and uses, commit to contemporary standards and policies, or try to use all types of assessment to both drive student learning and summate learning validly and reliably.

4.2. China
Teachers in China were less likely to be teacher-centric assessors than teachers in the U.S. but not Canada. Three times less than the expected number of teachers grouped with teacher-centric assessors, indicating that students in China are highly unlikely to encounter teacher-centric approaches, especially in comparison to their chances of encountering teacher-centric approaches in the U.S. However, teachers in China are generally proportionally represented in all other classes. In particular, more teachers than expected grouped with eager assessors and highly student-centric assessors, the latter of which showed no statistical differences across countries. These findings suggest that teachers in China respond to assessment macro- and micro-culture factors in a variety of ways. Students in China are likely to experience a range of different approaches to assessment that may vary highly from one teacher to the next. Assessment will be a significant part of their learning experience in some classes but not others. Some of their teachers will use assessment as a mechanism to improve learning, others will focus on measurement of learning, and still others will integrate assessment minimally and mobilize their efforts elsewhere.

4.3. Canada
Teachers in Canada were significantly more likely to be moderately student-centric assessors than teachers in the U.S. or China. They were also significantly more likely to be hesitant assessors than teachers in the U.S. but not China. Last, as no significant country differences were observed in highly student-centric assessors, teachers in Canada are equally likely to be highly student-centric assessors than teachers in the other two contexts.

These findings indicate that, in navigating assessment macro- and micro-cultures, teachers in Canada tend to adopt moderately student-centric, hesitant, or highly student-centric approaches to assessment. Classroom assessment will likely play a less central role in students' learning experiences compared to those of students in the U.S. However, both moderately and highly student-centric assessors value approaches that give students some control in assessment (e.g., AaL, differentiated), meaning that students in Canada are likely to experience more student-centered approaches designed to enhance metacognition and individual learning agency (e.g., student self-assessment, criteria co-construction, setting individual learning goals) than students in the U.S. or China.

5. Discussion
While previous research has looked at specific dimensions of classroom assessment (e.g., purposes, fairness, ethics) in the U.S., China, and Canada contexts separately, this study contributes
a cross-cultural, multidimensional view of teachers’ approaches to assessment as it varies between countries. Results from this study have uncovered important differences in how teachers approach classroom assessment and point to challenges students might face in adapting to new assessment macro- and micro-cultures. While our study did not aim to examine causal links between specific features of assessment macro- and micro-cultures, differing patterns of approaches both between and within countries support the notion that assessment macro- and micro-cultures influence experiences of assessment—and therefore learning—for students. Our study builds on scholarship that suggests that teachers’ understandings of assessment are shaped by sociocultural and policy traditions embedded in the fabric of society (Brown et al., 2019). As students and educators move between the U.S., China, and Canada, they will likely need to acculturate to different approaches to assessment shaped by policy and testing frameworks, as well as local school and classroom sociocultural knowledges.

This exploratory study offers initial insights into challenges that students and educators may face when moving between the U.S., China, and Canada. Specifically, when moving from China to the U.S., students will be moving from a highly accountability-driven assessment environment to one in which assessment is used widely across all levels of education, including formative and summative assessments, at classroom and large-scale levels. They will be integrating into a learning culture in which assessment is a high priority and engaging in a wide variety of assessment processes. However, the high prevalence of teacher-centric assessors builds on previous research that has suggested teachers in the U.S. use formative assessment primarily to guide their teaching (e.g., diagnostic quizzes, questioning; Barnes et al., 2017; Erickson, 2007). The observed teacher-centric approaches in the U.S. thus support findings that formative assessment strategies that build student agency and metacognition (e.g., self-assessment) are areas not widely realized in the U.S. context (Johnson et al., 2019). Our findings reflect the prevalence of a specific formative perspective on assessment in the U.S. that focuses on standardized diagnostic assessment rather than involving students actively in assessment processes (Bloom et al., 1971). While assessment will likely be a central part of students’ learning, the students themselves may have fewer opportunities to exercise agency in assessment processes and uses.

Based on our findings, students who move from China to Canada will likely face different acculturation challenges. Given the prevalence of moderately student-centric and hesitant assessors in Canada, our findings suggest that assessment will be a less central part of students’ learning experiences than in the U.S. However, if teachers in Canada are not hesitant assessors, they tend to be moderately or highly student-centric assessors, suggesting that teachers in Canada place high value on assessment approaches that support students’ metacognition and independent learning skills. The observed emphasis on AaL and individualized approaches does not reflect previous research on formative assessment practices in Canada (Volante, 2010) but does reflect policy frameworks (e.g., Ontario Ministry of Education, 2010) and more recent research examining pre-service teachers’ approaches to assessment (Coombs et al., 2020). One reason for the emphasis on AaL we observed in Canada could be that support for contemporary approaches (e.g., Klinger et al., 2015) among Canadian teachers has recently caught up with policy and theory. That teachers’ assessment approaches have caught up with contemporary standards is supported by more recent research in teacher candidate populations which also found greater endorsement of AaL (Coombs et al., 2020). However, given our relatively small sample size and that teachers in Canada are divided across many educational jurisdictions (i.e., provinces), future research is needed to verify this initial finding. Teachers’ emphasis on student-centric approaches in Canada might pose a particular challenge for students who move from China. Due to their high degree of familiarity with Chinese education, these students likely value teachers’ authority and expertise in the assessment process (Poole, 2016) and may be used to summative assessments that promote “transmission and memorization of ‘bookish’ knowledge for purely ranking or selection purposes” (Gan et al., 2017, p. 1126). While research has suggested the students from China are highly adaptable and strategic in their learning (Kennedy, 2016), assessment tasks that engage students in critical thinking about their own learning or permit student autonomy in assessment processes
might be a new experience for students from China who may require support in acculturating to new ways of thinking about assessment’s purpose.

6. Implications for practice and theory
This study has two important implications for practice. First, understanding how teachers navigate macro and micro-cultures in the U.S., China, and Canada provides a framework for supporting students who move between these contexts. Educators should be aware of the potential challenges faced by students moving from China to the U.S. and Canada and support students by explicitly discussing assessment macro- and micro-cultures. For example, a teacher in the U.S. could support a student entering a highly assessment-driven system by discussing approaches to assessment purposes, processes, fairness, and measurement to help prepare them for assessment situations. In contrast, teachers in Canada should focus on supporting students from China or the U.S. in navigating assessment processes designed to engage students’ metacognition and autonomy. Importantly, students entering any new assessment culture will need scaffolded learning about the assessment process in order to mitigate construct-irrelevant measurement effects and negative influences on assessment performance. In other words, teachers need to support newcomers in learning the assessment expectations, routines, and culture in similar ways to other schooling practices.

Second, professional development for teachers related to assessment should focus on the cultural dimensions of assessment in order to prepare teachers to understand and support students’ acculturation. Teachers should be prepared to manage both local and global understandings of assessment in an increasingly globalized world (DeLuca et al., 2020). Specifically, supporting students from abroad requires educators to reflect on how assessment macro- and micro-cultures shape their approaches to assessment.

This study also holds theoretical implications. While our findings reflect assessor-types identified in previous research (i.e., contemporary, eager, and hesitant; Coombs et al., 2020; Veldhuis & van den Heuvel-Panhuizen, 2014), two new classes emerged from our analysis: teacher-centric and moderately student-centric assessors. These new classes could be an artifact of assessment macro-cultures, teaching pressures, or our samples. As the teacher-centric assessor class emerged as a result of mainly U.S. participants’ patterns of endorsement, it is possible that accountability pressures that relate student outcomes on state tests to teacher performance (Committee on Incentives and Test-Based Accountability in Public Education, 2011) incentivize a focus on valid and reliable measurement, even in formative assessment. Teachers in the U.S. may be hesitant to prioritize student-centered, individualized approaches that may take time away from delivering content or informing teaching. The increasing accountability pressures U.S. teachers face could explain why our analysis uncovered teacher-centric approaches while other research has not. However, while standardized diagnostic assessment does not necessarily involve students actively in assessment processes, it is a long-recognized formative perspective on assessment (Bloom et al., 1971) and teachers in the U.S. may choose to adopt such approaches regardless of accountability pressures. The moderately student-centric assessor class, on the other hand, was largely driven by the Canada sample. These assessors relatively prioritize contemporary standards and policies (e.g., Klinger et al., 2015) but report only moderate endorsement of those approaches (e.g., AFL, equitable, balanced). An explanation for this previously undetected class is that moderately student-centric assessors espouse the approaches they learned in contemporary teacher education programs (DeLuca & Bellara, 2013; DeLuca & Klinger, 2010) but, given the need to manage other teaching and learning responsibilities (e.g., administrative tasks, lesson planning), they are more hesitant than highly student-centric assessors to endorse any response in classroom assessment scenarios. Balancing the many responsibilities of a full-time teacher might limit teachers’ capacity to take up various responses to assessment situations; this could explain why a recent latent class analysis of a pre-service teacher sample in Canada detected contemporary assessors (Coombs et al., 2020) but not moderately student-centric assessors. Without having
experienced the daily demands of full-time teaching, pre-service teachers may have idealistic responses to assessment scenarios.

Our research also adds support to the notion of assessment micro-cultures (Allal, 2016) as similar approaches were found across very different policy and testing contexts (i.e., assessment macro-cultures). Notably, highly student-centric assessors were proportionally represented in all three contexts. Notwithstanding significant associations between patterns of endorsement and country, that similar approaches were observed across these contexts highlights the role of local school and classroom social dynamics that comprise assessment micro-cultures. In other words, our findings reinforce that assessment macro-cultures do not shape teachers’ approaches alone; as similar approaches were observed across three highly different macro-cultures, assessment micro-culture elements (e.g., local knowledges, social dynamics, teacher beliefs) must be critical in shaping teachers’ approaches.

Last, our findings add nuance to Brown et al. (2019) assertion that education contexts with significant large-scale testing consequences discourage conceptions of assessment as a formative process. The teacher-centric approaches we found mainly in the U.S.—a context in which large-scale assessments have significant consequences for educators and students—highly endorsed AFL. However, their lack of support for AFL indicates that they see AFL primarily as a way to measure learning and inform teaching, rather than an opportunity to build students’ agency and metacognition. Thus, our study suggests that high-stakes testing cultures do not necessarily suppress AFL; rather, they shape a highly teacher-centric approach to it.

7. Future directions
There are limitations in this study that point to future directions for research. First, while the ACAI has been extensively utilized and validated (e.g., Coombs et al., 2020; DeLuca et al., 2016b), convergent validity evidence from classroom observations and interviews would add strength to the underlying assumption that self-reported assessment approaches reflect actual practice. As our study did not permit member checking, we cannot completely eliminate the possibility that participants that grouped in the highly student-centric, moderately student-centric, and eager assessors may have tried to respond to the ACAI scenarios in ways they perceived to be socially desirable. However, all possible scenario actions that could be taken by teachers are rooted in contemporary assessment policies (see DeLuca et al., 2016a) and are therefore reasonable and defensive actions a teacher could decide to employ. While this feature likely minimizes the effect of social desirability, observation data on teachers’ approaches to assessment in these three contexts could support more comprehensive understandings of these assessment cultures. Also, teachers who are interested in classroom assessment are more likely to complete the voluntary ACAI survey, meaning that moderately student-centric and hesitant assessor groups are likely less represented in the sample of this study than in the teaching population writ large. Research could utilize different methods, such as observations or surveying students about their teachers’ approaches, to explore the extent of less enthusiastic approaches to assessment. Given our findings, we hypothesize that hesitant and moderately student-centric approaches are associated with assessment as irrelevant conceptions of the purpose of assessment observed in the U.S. (Barnes et al., 2017) and China (Brown et al., 2011). Examining these relationships would build on understandings of how conceptions of assessment purposes translate to approaches. A fruitful line of research would also examine the relationships between relevant factors (e.g., teaching experience, teaching division) and hesitant or moderately student-centric approaches to uncover drivers of these approaches. Interviews with hesitant and moderately student-centric assessors could also illuminate why teachers hold these two approaches. Additionally, given that our findings support research that suggests teachers in the U.S. make limited use of student-centered formative assessment approaches (e.g., self- and peer assessment; Johnson et al., 2019), extensions of our study would interrogate the barriers that limit such practices. The suggested future directions would inform ways to support teachers in applying contemporary standards.
directions stemming from this study relate to testing causal assumptions about how macro- and micro-cultures shape teachers’ assessment approaches.

Second, our sample sizes are relatively small in proportion to the total population of teachers in each country. We wanted to investigate how the same classes appeared across the three contexts. As a result, we employed a LCA strategy with our whole sample to explore endorsement of the same five classes across the three country contexts; naturally, small numbers of teachers grouped into the classes that are likely unpopular in their particular context (e.g., few U.S. teachers grouped in the hesitant assessor class). However, the small number of teachers in some of the classes limits the generalizability of claims we can make about the prevalence of certain classes in the contexts. Thus, we have tried to present a balanced interpretation of our findings, the generalizability of claims about each country, and comparisons between contexts. Our study offers an initial exploration of the different approaches in the U.S., China, and Canada; findings describe a preliminary comparative perspective on the landscape of assessment across these contexts. Future research should aim to survey the approaches to assessment of more teachers in each context to enable more secure understandings of how assessment is approached in each context.

Third, the diversity of approaches observed in China in our study begins to add nuance to theories of washback effects on teaching and learning in test-driven contexts (e.g., teaching to the test; Chen & Brown, 2013; Yin & Buck, 2015). Rather than adopting primarily test-driven approaches (e.g., AoL, a standard fairness approach, consistent measurement), as previous washback research would suggest, the teachers we sampled navigate the Chinese assessment macro-culture, characterized by large-scale testing traditions, in various ways. Notably, many adopted highly student-centric approaches and most rejected teacher-centric ones. This finding suggests that the role of teacher autonomy and professional judgment may be more important in determining teachers’ approaches than testing traditions (Yung, 2002). Another explanation for our finding could be that policymakers’ efforts to mitigate the perceived negative impacts of high-stakes testing (Gan et al., 2017) are gaining traction among teachers in China. Future research should interrogate factors influencing teachers’ diverse approaches to assessment in China to build on understandings of washback theory. Additionally, research investigating the degree to which ACAI scenarios represent authentic and contemporary assessment situations in China would add clarity to the various approach patterns teachers endorsed. Interviewing participants who have taken the ACAI could elucidate how teachers in China interpret the items and help refine the ACAI scenarios. While Western conceptions of assessment have been adopted in Chinese policy, research has suggested that some principles do not translate in the Chinese sociocultural context (Poole, 2016). It is thus still important to understand assessment in its local context. Continued work to excavate local micro-cultures of assessment—and their cultural influences—remains at the heart of the assessment research agenda if we are to truly understand assessment from a sociocultural perspective.

8. Conclusion
These future directions highlight the initial steps our study has taken in comparing how teachers approach classroom assessment in the U.S., China, and Canada. Our study contributes empirical evidence that teachers’ approaches to assessment vary considerably across these contexts. Tapping into a multidimensional understanding of how teachers navigate and discern assessment macro- and micro-cultures, our findings describe differential patterns of assessment approaches across these contexts and have implications for supporting the acculturation of students who move between them. We contend that person-centered investigations of teachers’ approaches to assessment offer a unique opportunity to develop nuanced cross-cultural comparisons of how teachers understand and implement classroom assessment.
Funding
Social Sciences and Humanities Research Council of Canada, Grant #: 430-2013-0489

Author details
Christopher DeLuca
ORCID ID: http://orcid.org/0000-0001-5962-0827
Nathan Rickey
E-mail: nathan.rickey@queensu.ca
ORCID ID: http://orcid.org/0000-0002-0044-0204
Andrew Coombs
ORCID ID: http://orcid.org/0000-0002-4750-0964
1 Faculty of Education, Queen’s University, Kingston, Canada.

Citation information
Cite this article as: Exploring assessment across cultures: Teachers’ approaches to assessment in the U.S., China, and Canada, Christopher DeLuca, Nathan Rickey & Andrew Coombs, Cogent Education (2021), 8: 1921903.

References
Alberta Education. (2006). Effective student assessment and evaluation in the classroom. Edmonton, AB: Alberta Education. Retrieved from https://files.eric.ed.gov/fulltext/ED498247.pdf
Alberta Education. (2019). The guide to education: ECS to grade 12. Edmonton, AB: Alberta Education. Retrieved from https://open.alberta.ca/publications/1496-7359
Allal, L. (2015). The co-regulation of student learning in an assessment for learning culture. In D. Laveault & L. Allal (Eds.), Assessment for learning: Meeting the challenge of implementation (pp. 259–273). Springer.
Barnes, N., Fives, H., & Dacey, C. M. (2017). US teachers’ conceptions of the purposes of assessment. Teaching and Teacher Education, 65, 107–116. https://doi.org/10.1016/j.tate.2017.02.017
Barnes, N., Garie, C., DeLuca, C., Coombs, A. J., & Uchiyama, K. (2020). Exploring the roles of coursework and field experience in teacher candidates’ assessment literacy: A focus on approaches to assessment. Assessment Matters, 14, 5–41. doi:10.18296/am.0045
Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment: Implications for policy and practice. Policy Futures in Education, 13(1), 117–140. https://doi.org/10.1177/1478210314566733
Bloom, B., Hastings, J., & Madaus, G. (1971). Handbook on formative and summative evaluation of student learning. McGraw Hill.
British Columbia Ministry of Education. (2004). Classroom Assessment and Evaluation. BC Ministry of Education.
Brown, G. T., & Gao, L. (2019). Chinese teachers’ conceptions of assessment for and of learning: Six competing and complementary purposes. Cogent Education, 2(1), 2–19. https://doi.org/10.1080/2331186X.2014.993836
Brown, G. T., Gebril, A., & Michaelides, M. P. (2019). Teachers’ conceptions of assessment: A global phenomenon or a localism? Frontiers in Education, 4(16), 1–13. https://doi.org/10.3389/feduc.2019.00016
Brown, G. T., & 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 Multidisciplinary Evaluation, 6(12), 68–91.
Brown, G. T., Hui, S. K., Flora, 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., 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 & Practice, 16(3), 347–363. https://doi.org/10.1080/09695940802319737
Cai, Y., Wu, S., Zhou, L., & Zhong, M. (2017). An empirical study on teachers’ orientations of classroom evaluation purposes. Educational Measurement and Evaluation, 2, 20–32.
Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. Biometrics, 52(3), 195–212. https://doi.org/10.1007/BF02460998
Chen, J., & Brown, G. T. (2013). High-stakes examination preparation that controls teaching: Chinese preservice teachers’ conceptions of excellent teaching and assessment. Journal of Education for Teaching, 39(5), 541–556. https://doi.org/10.1080/02607476.2013.836338
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), 5. https://doi.org/10.5206/cie-eci.v47i1.9224
Cheng, L., Yan, W., Mei, Y., & DeLuca, C. (2017). Grading policies in China: are we assessing the learner or the learning? Assessment Matters, 11, 6–31. https://doi.org/10.18296/am.0022
Clark, S., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis [White paper]. Retrieved from https://www.statmodel.com/download/relatinglcga.pdf
Committee on Incentives and Test-Based Accountability in Public Education. (2011). In M. Hout & S. Elliott, Incentives and Test-Based Accountability in Education. Washington, D.C.: The National Academies Press. Retrieved from https://www.nap.edu/read/12521/chapter/1
Coombs, A., DeLuca, C., LoPoiteve-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: an empirical study across teacher career stages. Teaching and Teacher Education, 71, 134–146. https://doi.org/10.1016/j.tate.2017.12.010
Coombs, A., DeLuca, C., & MacGregor, S. (2020). A person-centered analysis of teacher candidates’ approaches to assessment. Teaching and Teacher Education, 87, 102952. https://doi.org/10.1016/j.tate.2019.102952
Coombs, A. J., Rickey, N. H., Liu, S., & DeLuca, C. (2021). Chinese teachers’ approaches to assessment: Educational Research for Policy and Practice. https://doi.org/10.1007/s10671-020-09289-z
Daniels, L. M., Poth, C., Papile, C., & Hutchison, M. (2016). Validating the conceptions of assessment-III scale in Canadian preservice teachers. Educational Assessment, 19(2), 139–158. https://doi.org/10.1080/10621719.2014.903624
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/10.1080/01446110.2017.1329380
DeLuca, C., & Bellora, A. (2013). The current state of assessment education: aligning policy, standards, and teacher education curriculum. Journal of Teacher
Education, 64(4), 356–372. https://doi.org/10.1177/0022487113488144

DeLuca, C., Braund, H., Vollquatte, A., & Cheng, L. (2017). Grading policies and practices in Canada: A landscape study. Canadian Journal of Educational Administration and Policy, 184, 4–22. https://jmsse.org/index.php/cjepa/article/view/16331

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. https://doi.org/10.1080/0969594X.2010.516643

DeLuca, C., LaPointe-McEwan, D., & Luhancho, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. Educational Assessment, 21(4), 248–266. https://doi.org/10.1080/10627199.2016.1236677

DeLuca, C., Vulliquette, A., Coombs, A., & LaPointe-McEwan, D., & Luhancho, U. (2018). Teachers’ approaches to classroom assessment: A large-scale survey. Assessment in Education: Principles, Policy & Practice, 25(4), 355–375. https://doi.org/10.1080/0969594X.2016.1244514

Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers’ grading and assessment practices. Alberta Journal of Educational Research, 53(1), 1–21. https://journalhosting.ucalgary.ca/index.php/ajer/article/view/55195

Erickson, F. (2007). Some thoughts on “Proximal” formative assessment of student learning. Yearbook of the National Society for the Study of Education, 106(1), 186–216. https://doi.org/10.1111/j.1744-7984.2007.00102.x

Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS). (2006). Definition of formative assessment. Council of Chief State School Officers.

Fulmer, G. W., Lee, J. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers’ assessment practices: an integrative review of research. Assessment in Education: Principles, Policy & Practice, 22(4), 475–494. https://doi.org/10.1080/0969594X.2015.1017445

Gan, Z., Liu, F., & Yang, C. C. R. (2017). Assessment for learning in the Chinese context: prospective EFL teachers’ perceptions and their relation to learning approach. Journal of Language Teaching and Research, 8(6), 1126–1134. https://doi.org/10.17507/jltr.0806.13

Gipps, C., Brown, M., McCullam, B., & McAlister, S. (1995). Intuition or evidence? Teachers’ and national assessment of seven-year-olds. Open University Press.

Gordon, E. W., McGill, M. V., Sands, D. I., Kalinich, K. M., Pellegrino, J. W., & Chatterji, M. (2014). Bringing formative classroom assessment to schools and making it count. Quality Assurance in Education, 22(4), 339–352

Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a foreign language in a Chinese secondary school? Assessment in Education: Principles, Policy & Practice, 21(3), 286–305. https://doi.org/10.1080/0969594X.2013.836076

Gu, Y. (2006). An ecological model of e-learning in Chinese context: critical reflections of 5 years’ practice of e-learning management in BOE. Studies in Continuing Education, 28(2), 99–120. https://doi.org/10.1080/01580370600750999

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2012). Standards-based accountability in the United States: Lessons learned and future directions. Education Inquiry, 3(2), 149–170. https://doi.org/10.3402/edui.v3i2.22025

Herppich, S., Praetorius, A., Förster, N., Glogger-Frey, I., Korst, 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.tea.2017.12.001

Hill, M. (2006). Dot, slash, cross: how assessment can drive teachers to tick instead of teaching. Set: Research Information for Teachers, (1), 21–25. https://doi.org/10.18296/set.0779

Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian english teachers’ practices for writing. Assessing Writing, 11(1), 42–65. https://doi.org/10.1016/j.awr.2005.12.002

Institute of International Education. (2019). Open doors: Report on international educational exchange. Retrieved from https://www.iie.org/Research-and-Insights/Open-Doors/Data/International-Students/Places-of-Origin.

Jin, L., & Cortazzi, M. (2006). Changing practices in Chinese cultures of learning. Language, Culture, and Curriculum, 19(1), 5–20. https://doi.org/10.1080/07908310608668751

Johnson, C. C., Sondergeld, T. A., & Walton, J. B. (2019). A study of the implementation of formative assessment in three large Urban Districts. American Educational Research Journal, 56(6), 2408–2438. https://doi.org/10.3102/0021965419837524

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. Sammons & L. R. Harris (Eds.), Handbook of human and social conditions in assessment (pp. 304–419). Routledge.

Kim, D. H., Lambert, R. G., & Burts, D. C. (2018). Are young dual language learners homogeneous? Identifying subgroups using latent class analysis. The Journal of Educational Research, 111(1), 43–57. https://doi.org/10.1080/00220671.2016.1190912

Klinger, D., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. Canadian Journal of Educational Administration and Policy, 76, 1–34. https://cdm.ucalgary.ca/index.php/cjepa/article/view/42757

Klinger, D. A., McDevitt, P. R., Howard, B. B., Munoz, M. A., Rogers, W. T., & Wylie, E. C. (2015). The Classroom Assessment Standards for PreK-12 Teachers. Kindle Direct Press

Klinger, D. A., & Rogers, W. T. (2011). Teachers’ perceptions of large-scale assessment programs within low-stakes accountability frameworks. International Journal of Testing, 11(2), 122–143. https://doi.org/10.1080/15305058.2011.552748

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class
analysis. Structural Equation Modeling: A Multidisciplinary Journal, 14(4), 671–694. https://doi.org/10.1080/2331186X.2021.1921903

Lashway, L. (2001). The New Standards and Accountability: Will Rewards and Sanctions Motivate America’s Schools to Peak Performance?. ERIC Clearinghouse on Educational Management.

Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: implications for assessment literacy. Assessment in Education: Principles, Policy & Practice, 17(1), 7–21. https://doi.org/10.1080/09695940903565362

Li, Y. (2004). Learning to live and study in Canada: stories of four EFL learners from China. TESL Canada Journal, 22(1), 25–43. https://doi.org/10.18868/tesl.v22i1.164

Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers’ perceptions about ethical issues in classroom assessment. Studies in Educational Evaluation, 48, 57–66. https://doi.org/10.1016/j.stueduc.2016.01.002

Lo, Y., Mendell, N. R., & Rubin, D. B. (2000). Testing the number of components in a normal mixture. Biometrika, 88(3), 767–778. https://doi.org/10.1093/biomet/88.3.767

McMillan, J. (2013). Classroom assessment in the context of learning theory and research. In J. McMillan (Ed.), SAGE handbook of research on classroom assessment (pp. 1–16). Sage.

McMillan, J., Myran, S., & Workman, D. (2003). Elementary teachers’ classroom assessment and grading practices. Journal of Educational Research, 95(4), 203–213. https://doi.org/10.1080/00220670209596593

Muthén, B., & Muthén, L. K. (2015). Mplus version 8. Retrieved from https://www.statmodel.com

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. Alcoholism, Clinical and Experimental Research, 24(6), 882–891. https://doi.org/10.1111/j.1530-0277.2000.tb02070.x

National Research Council. (2001). Knowing what students know: The science and design of educational assessment. In J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), Committee on the Foundations of Assessment (pp. 1–16). Board on Testing and Assessment, Center for Education, Division ofBehavioral and Social Sciences and Education, National Academies Press.

Nichols, S. L., & Harris, L. R. (2016). Accountability assessment’s effects on teachers and schools. In G. T. L. Brown & L. R. Harris (Eds.), Handbook of human and social conditions in assessment (pp. 40–56). Routledge.

Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: what does latent class analysis say?. Child Development, 78(6), 1706–1722. https://doi.org/10.1111/j.1467-8624.2007.01097.x

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14(4), 535–569. https://doi.org/10.1080/10705510701575396

O’Keefe, B., & Lewis, B. (2019). The state of assessment: A look forward on innovation in testing systems. Bellwether Education Partners. Retrieved from https://files.eric.ed.gov/fulltext/ED596503.pdf

OECD. (2011). Strong performers and successful reformers in education: Lessons from PISA for the United States. Ontario Ministry of Education. (2010). Growing success: Assessment, evaluation and reporting in Ontario classrooms. Queen’s Printer for Ontario.

Perie, M., Marion, S., & Gong, B. (2007). A framework for considering interim assessments. National Center for the Improvement of Educational Assessments.

Peterson, S. S., & McClay, J. (2010). Assessing and providing feedback for student writing in Canadian classrooms. Assessing Writing, 15(2), 86–99. https://doi.org/10.1016/j.asw.2010.05.003

Poole, A. (2016). ‘Complex teaching realities’ and ‘deep rooted cultural traditions’: barriers to the implementation and internalisation of formative assessment in China. Cogent Education, 3(1), 1156242. https://doi.org/10.1080/2331186X.2016.1156242

Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. Teaching and Teacher Education, 25(5), 778–782. https://doi.org/10.1016/j.tate.2008.11.013

Roesch, S. C., Viladoms, M., & Viladoms, F. (2010). Latent class/profile analysis in maltreatment research: A comment on Noelle et al., Pears et al., and looking beyond. Child Abuse & Neglect, 34(3), 155–160. https://doi.org/10.1016/j.chiabu.2010.01.003

Shepard, L. A. (2005, October). Formative assessment: caveat emptor. Paper presentation at the ETS invitation conference the future of assessment: shaping teaching and learning, New York, NY.

Shepard, L. A. (2000). The role of classroom assessment in teaching and learning (Report No. 517). Center for the Study of Evaluation. The Regents of the University of California. Retrieved from https://nepc.colorado.edu/sites/default/files/publications/TECH517.pdf

Shepard, L. A. (2009). Commentary: evaluating the validity of formative and interim assessment. Educational Measurement: Issues and Practice, 28(3), 32–37. https://doi.org/10.1111/j.1745-3992.2009.00152.x

Statistics Canada. (2016). International students in Canadian universities, 2004/2005 to 2013/2014. Retrieved from http://www.statcan.gc.ca/pub/81-599-x/81-599-x2016011-eng.pdf

Stiggins, R. J. (2002). Assessment crisis: the absence of assessment for learning. Phi Delta Kappan, 83(10), 758–765. https://doi.org/10.1177/003172430208301020

Stobart, G., & Eggen, T. (2012). High-stakes testing–value, fairness and consequences. Assessment in Education: Principles, Policy & Practice, 19(1), 1–6. https://doi.org/10.1080/09695954.2012.639191

Tota, J. (2005). The influence of national culture on the perceived fairness of grading procedures: A comparison of the United States and China. Journal of Psychology, 139(5), 401–412. https://doi.org/10.3200/JRLP.139.5.401-412

The Gordon Commission. (2013). Assessment for learning: A public policy statement. Retrieved from http://www.gordoncommission.org/resources/pdfs/gordon_commission_public_policy_report.pdf

The Ministry of Education of P. R. China. (2012). National professional standards for K12 teachers. Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/s6991/201212/xxgk_145603.html

Tierney, R. D. (2006). Changing practices: influences on classroom assessment. Assessment in Education, 13(3), 239–264. https://doi.org/10.1080/09695950601035387

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. Studies in Educational
Evaluation, 43, 55–69. https://doi.org/10.1016/j.stueduc.2013.12.003
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
Timmons, K., & Pelletier, J. (2016). Using latent-class analysis to examine the influence of kindergarten children’s perspectives of school on literacy and self-regulation outcomes. Early Child Development and Care, 186(11), 1737–1751. https://doi.org/10.1080/03004430.2015.1311577
US Department of Education. (n. d.). Every student succeeds act (ESSA). US Department of Education. Retrieved from https://www2.ed.gov/essa?src=policy
Veldhuis, M., & Van Den Heuvel-panhuizen, M. (2014). Primary school teachers’ assessment profiles in mathematics education. PloS One, 9(1), 1–11. https://doi.org/10.1371/journal.pone.0086817
Volante, L. (2010). Assessment of, for, and as learning within schools: implications for transforming classroom practice. Action in Teacher Education, 31(4), 66–75. https://doi.org/10.1080/01626620.2010.10463536
Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: synergies and tensions between research and practice. Canadian Journal of Education, 34(2), 239–255. https://www.jstor.org/stable/10.2307/canajeducrevcan.34.2.239
Volante, L., & Jafdar, B. (2008). Educational assessment in Canada. Assessment in Education: Principles, Policy & Practice, 15(2), 201–210. https://doi.org/10.1080/09695940802164226
Wang, J. (2012). Curriculum reform in mainland China, 1978–2008: change, maintenance, and conflicts. Chinese Education & Society, 45(1), 59–68. https://doi.org/10.2753/CED1061-1932450107
Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers’ assessment literacies in an era of curriculum and assessment reform. The Australian Educational Researcher, 40(2), 241–256. https://doi.org/10.1007/s13384-013-0089-9
Xu, W. (2004). Society modality, politics power and education system: the logic of modern Chinese education system reform. Fudan Education Forum, 2(4), 21–25.
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
Xu, Y., & Brown, G. T. L. (2017). University english teacher assessment literacy: A survey-test report from China.

Papers in Language Testing and Assessment, 6(1), 133–158. http://www.altanza.org/uploads/59/9/08/5908292/9/si7xubrown_final_formatted_proofed.pdf
Yin, X., & Buck, G. A. (2015). There is another choice: An exploration of integrating formative assessment in a Chinese high school chemistry classroom through collaborative action research. Cultural Studies of Science Education, 10(3), 719–752. https://doi.org/10.1007/s11422-014-9572-5
Yung, B. H. W. (2002). Same assessment, different practice: professional consciousness as a determinant of teachers’ practice in a school-based assessment scheme. Assessment in Education: Principles, Policy & Practice, 9(1), 97–117. https://doi.org/10.1080/09695940202192120
Zhang, D., Ding, Y., Lee, S., & Chen, J. (2017). Strategic development of multiplication problem solving: patterns of students’ strategy choices. The Journal of Educational Research, 110(2), 159–170. https://doi.org/10.1080/00220671.2015.1060928
Zhang, Z., & Zhou, G. (2010). Understanding Chinese international students at a Canadian university: perspectives, expectations, and experiences. Comparative and International Education/éducation comparée Et Internationale, 39(3), 43–58. https://doi.org/10.5206/cie-eci.v39i3.9162
Zhao, X., Van Den Heuvel-panhuizen, M., & Veldhuis, M. (2016). Teachers’ use of classroom assessment techniques in primary mathematics education—an explorative study with six Chinese teachers. International Journal of STEM Education, 3(1), 1–18. https://doi.org/10.1186/s40594-016-0051-2
Zhao, X., Van Den Heuvel-panhuizen, M., & Veldhuis, M. (2017). Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers. Studies in Educational Evaluation, 52, 42–54. https://doi.org/10.1016/j.stueduc.2016.12.002
Zheng, D. (2017). Strategies for improvement of teacher assessment literacy. People’s Education, 15, 33–38.
Zhihua, Z., & Kumari, B. (2014). I came, but I’m lost: learning stories of three Chinese international students in Canada. Comparative and International Education/éducation comparée Et Internationale, 43 (2), 1–14. https://doi.org/10.5206/cie-eci.v43i2.9252
Zhou, G., & Zhang, Z. (2014). A study of the first year international students at a Canadian university: challenges and experiences with social integration. Canadian and International Education/éducation Canadienne Et International, 43(2), 7. https://doi.org/10.5206/cie-eci.v43i2.9253
