

Teacher assessment literacy: a review of international standards and measures

Christopher DeLuca¹ ·
Danielle LaPointe-McEwan¹ · Ulemu Luhanga¹

Received: 1 October 2014 / Accepted: 18 November 2015 / Published online: 27 November 2015
© Springer Science+Business Media New York 2015

Abstract Assessment literacy is a core professional requirement across educational systems. Hence, measuring and supporting teachers' assessment literacy have been a primary focus over the past two decades. At present, there are a multitude of assessment standards across the world and numerous assessment literacy measures that represent different conceptions of assessment literacy. The purpose of this research is to (a) analyze assessment literacy standards from five English-speaking countries (i.e., Australia, Canada, New Zealand, UK, and USA) plus mainland Europe to understand shifts in the assessment landscape over time and across regions and (b) analyze prominent assessment literacy measures developed after 1990. Through a thematic analysis of 15 assessment standards and an examination of eight assessment literacy measures, results indicate noticeable shifts in standards over time yet the majority of measures continue to be based on early conceptions of assessment literacy. Results also serve to define the multiple dimensions of assessment literacy and yield important recommendations for measuring teacher assessment literacy.

Keywords Assessment · Measurement · Assessment literacy · Teacher knowledge · Assessment competency · Assessment standards

1 Introduction

Teacher assessment literacy (i.e., teacher competency in educational assessment) is a professional requirement within the current accountability framework of public education across many parts of the world (DeLuca 2012; Popham 2013; Volante and Fazio 2007). Assessment literacy involves the ability to construct reliable assessments and

✉ Christopher DeLuca
cdeluca@queensu.ca

¹ Faculty of Education, Queen's University, A218 Duncan McArthur Hall, 511 Union Street, Kingston, ON, Canada K7M 5R7

then administer and score these assessments to facilitate valid instructional decisions anchored to state or provincial educational standards (Popham 2004, 2013; Stiggins 2002, 2004). Recent policy developments throughout North America, Europe, Australia, and New Zealand have emphasized classroom teachers' ongoing formative and summative assessments to guide instruction and support student learning (Birenbaum et al. 2015). Further, empirical studies have demonstrated significant gains in student achievement, metacognitive functions, and motivation for learning when teachers integrate assessment with their instruction (Black and Wiliam 1998; Earl 2003; Gardner 2006; Willis 2010). Despite these potential benefits, research continues to show that teachers struggle to interpret assessment policies and to implement assessment practice in alignment with contemporary mandates and assessment theories (DeLuca and Klinger 2010; MacLellan 2004). Moreover, researchers have noted that there is comparatively little research on teachers' current assessment practices from which to construct responsive professional learning structures aimed at promoting teacher assessment literacy (Mertler 2009).

One predominant reason for a lack of reliable research on teachers' assessment literacy is that "the psychometric evidence available to support assessment literacy measures is weak" (Gotch and French 2014, p. 16). Through their recent systematic review of the psychometric properties of 36 assessment literacy measures, Gotch and French found that despite assessment literacy being a national priority in the USA and a keystone component of teacher evaluation, existing measures maintain weak evidence across reliability and validity indicators of test content, internal consistency reliability, score stability, and association with student outcomes. Gotch and French conclude that in order to increase the validity of assessment literacy measures researchers must begin by examining the "representativeness and relevance of content in light of transformations in the assessment landscape (e.g., accountability systems, conceptions of formative assessment)" (p. 17).

Specifically, and following Brookhart (2011), they argue that measures need to be constructed and analyzed in relation to contemporary assessment standards, reflective of the current context for teacher assessment practice, and must move beyond their dominant reference to the 1990 *Standards for Teacher Competency in Educational Assessment of Students* (AFT et al. 1990). Brookhart (2011) noted that the 1990 Standards have become dated in two ways: (a) they do not consider current conceptions of formative assessment (i.e., assessment for learning), and (b) they do not consider the technical and social issues teachers face in constructing and using assessments within standards-based educational reforms. Accordingly, Brookhart identified revisions to the standards in order to appropriate them for the current accountability context of education. These revisions align with the soon-to-be published *Classroom Assessment Standards* (Joint Committee on Standards for Educational Evaluation 2015), a set of research-based principles and guidelines for assessing student learning in classroom contexts.

In light of critiques directed at previous assessment literacy measures and based on recommendations to develop instruments based on contemporary assessment standards and practices (Brookhart 2011; Gotch and French 2014), this paper reviews current assessment standards and existing assessment literacy measures from selected English-speaking countries and mainland Europe. These countries and regions were selected because of their longstanding representation at the International Symposium for

Classroom Assessment (ISCA *n.d.*) and persistent commitment to the development of assessment theory, policy, and practice. Specifically, these regions are Australia, Canada, New Zealand, UK, USA, and mainland Europe. Hence, the purposes of this research are to (a) analyze assessment literacy standards from six regions (i.e., Australia, Canada, New Zealand, UK, USA, and mainland Europe) to understand shifts in the assessment landscape over time and across regions and (b) analyze prominent assessment literacy measures developed post-1990 Standards. Given the pervasiveness of the accountability movement across current systems of education, our fundamental aim in conducting this research is to provide a starting point for future development of assessment literacy measures that more accurately align with teachers' current assessment demands.

2 Methods

A two-phase research design was used to achieve the dual purposes of this study. The first phase involved collecting and analyzing documents describing teacher assessment literacy standards from various regions: Australia, Canada, New Zealand, UK, USA, and mainland Europe. To build on Gotch and French's (2014) review, the second phase involved examining post-1990 assessment literacy measures to analyze the degree to which these measures align with contemporary assessment standards.

2.1 Phase 1: analysis of assessment literacy standards

Assessment standards were selected from the six English-speaking regions (i.e., Australia, Canada, New Zealand, UK, USA, and mainland Europe). These regions have a demonstrated commitment to the advancement of classroom assessment research, policy, and practice as evident by their longstanding representation (since 2001) at the International Symposium for Classroom Assessment (ISCA *n.d.*) and with assessment leaders within the Consortium of International Researchers in Classroom Assessment. These countries have dedicated "standards" documents that shape and guide teacher practice in the area of classroom assessment. That said, it is important to note that other countries and regions have strong traditions in classroom assessment with delineated policies for teacher practice (both English-speaking and not). Hence, the regional selection criteria for this study while justified do represent a generalizability limitation with this research. Therefore, we assert this study as a starting point for future research on assessment standards for teacher practice.

Within the selected countries and regions, assessment standards were systematically identified by reviewing: (a) public websites for national or inter-state organizations and governmental ministries of education (e.g., Australian Department of Education, Department of Education in the United Kingdom, and US Department of Education); (b) national or inter-state assessment research consortia, associations, and joint advisory committees (e.g., Assessment and Certification Authorities, Assessment Reform Group, Association for Educational Assessment-Europe, Australian Curriculum, Joint Advisory Committee-Canada, National Council on Measurement in Education, and Joint Committee for Standards on Educational Evaluation); and (c) national and regional teacher education associations (e.g., Interstate Teacher Assessment and

Support Consortium, National Board for Professional Teaching Standards, and National Council for the Accreditation of Teacher Education). Only documents that explicitly addressed “standards” for teacher competency or literacy in assessment of student learning were selected for further analysis; in total, 15 standards documents were identified (see Table 1). Only national or inter-state level policies were used for this analysis. In countries with decentralized educational systems in which education falls within state or provincial jurisdiction (e.g., Canada and the USA), there may be additional policies that operate at state/provincial levels.

All 15 documents were first coded by region and date of publication. Documents were then analyzed inductively using standard thematic coding procedures (Patton 2002) (see Table 2). The unit of analysis for thematic coding was each standard or, where applicable, its associated guidelines. For each document, frequencies were constructed to show the representation of each identified code. The total frequency of

Table 1 Assessment standards document descriptions

Document	Origin	Description
Australian Professional Standards for Teachers (Australia Dept. of Education 2012)	Australia	7 standard and related guidelines
Brookhart (2011)	USA	11 standards
Changing Assessment Practices: Process, Principles and Standards (ARG 2008)	UK	4 standards and related guidelines
Classroom Assessment Standards: Practices for PK-12 Teachers (JCSEE 2015)	USA	16 standards and related guidelines
European Framework of Standards for Educational Assessment 1.0 (AEA-Europe 2012)	Europe	7 core elements
Graduating Teacher Standards (New Zealand Teacher New Zealand Teachers Council 2008)	NZ	7 standards and related guidelines
Guidelines for Assessment Quality and Equity (ACACA 1995)	Australia	20 guidelines
InTASC Model Core Teaching Standards: A Resource for State Dialogue (InTASC 2011)	USA	10 standards
Principles for Fair Student Assessment Practices for Education in Canada (JAC 1993)	Canada	5 standards and related guidelines
Professional Responsibilities in Measurement (NCME 1995)	USA	8 standards and related guidelines
Professional Standards for the Accreditation of Teacher Preparation Institutions (NCATE 2008)	UK	6 standards and specific program standards
Revised Teacher Standards (UK Department of Education 2012)	UK	8 standards
Standards for Educational and Psychological Testing (AERA et al. 1999)	USA	16 standards
The Standards for Teacher Competence in Educational Assessment of Students (AFT et al. 1990)	USA	7 standards
What Teachers Should Know and Be Able to Do (NBPTS 2001)	USA	5 core propositions

Table 2 Assessment standards theme descriptions and code criteria

Theme	Associated codes	Description of theme
Assessment purposes	<ul style="list-style-type: none"> ▪ MULTIPLE forms of assessment ▪ FORMATIVE vs. summative assessment ▪ Classroom-based vs. standardized assessment 	Choosing the appropriate form of assessment based on clearly stated instructional goals.
Assessment processes	<ul style="list-style-type: none"> ▪ Developing assessments ▪ Administering assessments ▪ Collecting assessment information ▪ Scoring assessments ▪ Interpreting scores ▪ Using assessments to make decisions ▪ Developing valid student grading procedures ▪ Monitoring and revising assessment processes 	Constructing, administering, and scoring assessments. Interpreting assessment results to facilitate instructional decision-making.
Communication of assessment results	<ul style="list-style-type: none"> ▪ Communicating assessment purposes and processes ▪ Articulating grading procedures ▪ Communicating results to parents and stakeholders 	Communicating assessment purposes, processes, and results to stakeholders.
Assessment fairness	<ul style="list-style-type: none"> ▪ Employing fair assessment practice ▪ Providing fair representation of students' achievement ▪ Accommodating exceptional learners ▪ Acknowledging diversity 	Cultivating fair assessment conditions for all learners with sensitivity to student diversity and exceptional learners.
Assessment ethics	<ul style="list-style-type: none"> ▪ Disclosing accurate, balanced information ▪ Protecting rights and privacy ▪ Minimizing biases ▪ Complying with standards 	Disclosing accurate information about assessments. Protecting the rights and privacy of students that are assessed.
Measurement Theory	<ul style="list-style-type: none"> ▪ Reliability ▪ Validity ▪ Large-scale assessments ▪ Norms and standards 	Understanding psychometric properties of assessments (e.g., reliability and validity).
Assessment for learning	<ul style="list-style-type: none"> ▪ Formative assessment ▪ Diagnostic assessment ▪ Assessment for learning ▪ Student self- and peer-assessment ▪ Feedback to students ▪ Student engagement in assessment practices 	Using formative assessment during instruction to guide teacher practice and student learning.
Education and support for teachers	<ul style="list-style-type: none"> ▪ Educating users of large-scale assessments ▪ Providing opportunities for teachers to develop assessment competency ▪ Supporting teachers' assessment practices and competency 	Educating and supporting teachers' assessment competency.

a code in relation to the total number of standards or guidelines within the document was calculated and expressed as a percentage. This proportion-based process reduced the inflation of frequency counts across documents with varying numbers of standards and/or guidelines (DeLuca and Bellara 2013). Codes were then collapsed into themes, and total percentages for each theme were reported (Tables 3 and 4). In total, we identified 35 codes that were collapsed into eight themes. These themes were (a) Assessment Purposes, (b) Assessment Processes, (c) Communication of Assessment Results, (d) Assessment Fairness, (e) Assessment Ethics, (f) Measurement Theory, (g)

Table 3 Theme frequencies (expressed as percentages) for assessment standards documents from USA and Canada

Theme	Government and research-based assessment standards						Teacher accreditation- and certification-based standards		
	AFT et al. (1990)	JAC (1993)	NCME (1995)	AERA (1999)	Brookhart (2011)	JCSEE (2015)	NBPTS (2001)	NCATE (2008)	InTASC (2011)
Assessment purposes	14.3	25.0	3.5	31.2	18.2	17.6	2.9	0.0	3.4
Assessment processes	57.1	58.3	12.8	12.5	18.2	29.5	0.0	7.1	3.4
Communication of assessment results	14.3	16.7	3.5	6.3	9.0	5.9	2.9	0.0	0.0
Assessment fairness	14.3	0.0	22.0	25.0	9.0	23.5	0.0	0.0	2.9
Assessment ethics	0.0	0.0	53.5	0.0	0.0	0.0	0.0	0.0	0.0
Measurement Theory	0.0	0.0	2.4	25.0	0.0	0.0	0.0	0.0	0.0
Assessment for learning	0.0	0.0	0.0	0.0	27.4	17.6	8.8	7.1	8.6
Assessment education and support for teachers	0.0	0.0	2.3	0.0	18.2	5.9	0.0	7.1	0.6

Assessment for Learning, and (h) Assessment Education and Support for Teachers. Themes and their associated codes are described in Table 2. Following contemporary document analysis methods, all documents were coded by two raters (Bowen 2009; Patton 2002); any disagreements in coding were discussed to reach consensus. Results from this phase were further analyzed in relation to (a) shifts in assessment standards over time and (b) variations in assessment standards between regions.

Table 4 Theme frequencies (expressed as percentages) for assessment standards documents from Australia, New Zealand, UK, and mainland Europe

Theme	Government and research-based assessment standards			Teacher accreditation- and certification-based standards		
	ACACA (1995)	ARG (2008)	AEA-Europe (2012)	NZ (2008)	Australia (2012)	UK (2012)
Assessment purposes	10.0	10.9	14.3	3.4	5.4	2.6
Assessment processes	35.0	23.9	71.4	0.0	2.7	2.6
Communication of assessment results	0.0	10.9	14.3	3.4	2.7	2.6
Assessment fairness	35.0	2.1	0.0	0.0	0.0	0.0
Assessment ethics	0.0	0.0	0.0	0.0	0.0	0.0
Measurement Theory	5.0	0.0	0.0	0.0	0.0	0.0
Assessment for learning	0.0	41.3	0.0	3.4	8.1	7.9
Assessment education and support for teachers	15.0	10.9	0.0	0.0	0.0	2.6

2.2 Phase 2: analysis of existing assessment literacy measures

In order to identify contemporary, post-1990 assessment measures, a systematic review of empirical studies related to assessment literacy was conducted. ERIC, PsycINFO, and ProQuest Dissertation and Theses databases were searched using the term *assessment literacy*. The search was restricted to published English language instruments that examined assessment literacy or teacher competency in assessment for pre-service and in-service teachers. Eight instruments published between 1993 and 2012 were identified: Assessment Literacy Inventory; Assessment Practices Inventory; Assessment Self-Confidence Survey; Assessment in Vocational Classroom Questionnaire, Part II; Classroom Assessment Literacy Inventory; Measurement Literacy Questionnaire; revised Assessment Literacy Inventory; and the Teacher Assessment Literacy Questionnaire (see Table 5).

We used a two-phase process to analyze the eight assessment instruments. First, we analyzed each instrument based on (a) its item characteristics (i.e., number of content based items, Likert-type items, scenario based items, and true/false items), (b) the instrument's guiding framework (i.e., standards or literature used for instrument blueprint); and (c) the instrument's reported psychometric properties. Second, we deductively analyzed instrument items in relation to the eight identified themes from phase 1 of the research. Items in which more than one theme was represented were dual coded (i.e., primary and secondary themes), and code disagreements were discussed until consensus was reached. The total frequency of a theme in relation to the total number of items per instrument was then calculated and expressed as a percentage (see Table 6).

3 Results

3.1 Assessment literacy: what is it?

The results from phase 1 provide insight into standards that delineate teacher assessment literacy. Specifically, nine governmental and research-based assessment standards documents were included in our thematic analysis: five from the USA and one each from Canada, Australia, the UK, and mainland Europe. In addition, six standard documents from teacher accreditation and certification organizations were included; three from the USA and one each from New Zealand, Australia, and the UK. Prior to analyzing temporal and regional variations in these various documents, we briefly describe the number of standards and guidelines found within each document (see Table 1).

3.2 Governmental and research-based assessment standards

In the USA, the *Standards for Teacher Competence in Educational Assessment of Students* (AFT et al. 1990) document was created to guide teacher educators and teachers in developing assessment competency. The document comprises seven standards that have been widely represented and reinforced in assessment textbooks for teachers, teacher education courses, policy documents, and educational research (Brookhart 2011).

Table 5 Assessment literacy instrument descriptions

Instrument	Item characteristics	Guiding framework	Respondents	Psychometric properties
Assessment Literacy Inventory (ALI) (Campbell et al. 2002)	35 content-based items (5 items per standard)	1990 Standards	220 pre-service teachers	$\alpha=0.74$; mean total score=21
Assessment Practices Inventory (API) (Zhang and Burry-stock 1997)	67 Likert-type items; 5-point scale (1=Not at all skilled to 5=Highly skilled)	1990 Standards	311 in-service teachers	$\alpha=0.97$ for total instrument; α ranged from 0.79 to 0.93 for 7 subscales
Assessment Self-Confidence Survey (Jarr 2012)	15 Likert-type items; 7-point scale (1=Not confident at all to 7=Very confident)	Bandura's (2006) guidelines for constructing self-efficacy scales	201 in-service teachers	$\alpha=0.90$; mean total score=64.9 (out of 105 max. points), SD 14.2
Assessment in Vocational Classroom Questionnaire, Part II (Kershaw IV 1993)	26 Likert-type items; 5-point scale (1=Not competent to 5=Extremely competent)	1990 Standards	393 in-service teachers	$\alpha=0.91$; mean total score=97.0 (out of 130 max. points), SD=12.9
Classroom Assessment Literacy Inventory (CALI) (Mertler 2003)	35 content-based items (5 items per standard)	1990 Standards	197 in-service teachers	$\alpha=0.57$; mean total score=22.0, SD=3.4
Measurement Literacy Questionnaire (Daniel and King 1998)	30 True/False items	Assessment literature (e.g., Gullickson 1984; Kubiszyn and Borich 1996; Popham 1995)	67 pre-service teachers	$\alpha=0.74$; mean total score=19.0, SD=4.7
Revised Assessment Literacy Inventory (ALI) (Mertler and Campbell 2005)	35 scenario-based items (5 scenarios; 5 items per standard)	1990 Standards	95 in-service teachers	$\alpha=0.60$; mean total score=18.2, SD=3.3
Teacher Assessment Literacy Questionnaire (TALQ) (Plake et al. 1993)	35 content-based items (5 items per standard)	1990 Standards	250 pre-service teachers	$\alpha=0.74$; mean total score=23.9, SD=4.6
			555 in-service teachers	$\alpha=0.54$; mean total score=23.2, SD=3.3

1990 Standards=Standards for Teacher Competence in the Educational Assessment of Students (AFT et al. 1990)

Table 6 Mapping assessment literacy instrument items onto assessment standards themes

	ALI	API	ASCS	AVCQ-II	CALI	MLQ	Revised ALI	TALQ
Themes No. of items	35	67	15	26	35	30	35	35
Assessment purposes	14.3 %	6.0 %	0.0 %	11.5 %	14.3 %	3.3 %	14.3 %	14.3 %
Assessment processes	57.1 %	58.2 %	46.7 %	61.5 %	57.1 %	53.3 %	57.1 %	57.1 %
Communication of assessment results	14.3 %	10.4 %	26.7 %	11.5 %	14.3 %	0.0 %	14.3 %	14.3 %
Assessment fairness	0.0 %	11.9 %	6.7 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
Assessment ethics	14.3 %	4.5 %	6.7 %	3.8 %	14.3 %	0.0 %	14.3 %	14.3 %
Measurement Theory	0.0 %	6.0 %	0.0 %	11.5 %	0.0 %	43.3 %	0.0 %	0.0 %
Assessment for learning	0.0 %	3.0 %	13.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
Assessment education and support for teachers	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %

ASCS Assessment Self-Confidence Survey, *AVCQ-II* Assessment in Vocational Classroom Questionnaire, Part II (Competence in Assessment), *MLQ* Measurement Literacy Questionnaire

Three years later, the *Principles for Fair Student Assessment Practices for Education in Canada* (Joint Advisory Committee 1993) were published in Canada. This document consists of five standards and related guidelines intended to ensure fair assessment practices in Canadian educational contexts. The document is aimed at users and developers of classroom-based and standardized assessments. The document was generated from a cross-Canadian panel with two representatives appointed from nine Canadian educational organizations (i.e., Canadian Education Association, Canadian School Boards Association, Canadian Association for School Administrators, Canadian Teachers Federation, Canadian Guidance and Counselling Association, Canadian Association of School Psychologists, Canadian Council for Exceptional Children, Canadian Psychological Association, and Canadian Society for the Study of Education).

Back in the USA, the National Council on Measurement in Education (NCME) produced the *Code of Professional Responsibilities in Measurement* in 1995. This extensive document was intended to guide all individuals involved in educational assessment activities, formal and informal, to “uphold the integrity of the manner in which assessment are developed, used, evaluated, and marketed” (p. 1). The NCME document comprises eight standards with associated guidelines.

In 1999, the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education published the *Standards for Educational and Psychological Testing*. These sixteen standards serve as a reference for professional test developers, policy makers, and test users in the domains of education, psychology, and employment. A revised version is anticipated in 2014.

Recognizing the limitations of previous assessment standards documents, Brookhart (2011) proposed a new set of educational assessment knowledge and skills for teachers. Specifically, Brookhart argued that the *Standards for Teacher Competence in Educational Assessment of Students* (AFT et al. 1990) failed to incorporate two

significant developments in educational assessment: (a) formative assessment and (b) standards-based reform. Consequently, she proposed a set of 11 standards that reflect teachers' current assessment competency needs.

Most recently, the Joint Committee for Standards on Educational Evaluation (2015) released the *Classroom Assessment Standards: Practices for PK-12 Teachers*. This document includes 16 standards and related guidelines that illustrate essential considerations when “exercising the professional judgment required for fair and equitable classroom formative, benchmark, and summative assessments for all students” (p. 1). These standards can be used by teachers, students, and parents/guardians to support and enhance student learning.

At the other ends of the globe, the Australian Curriculum, Assessment and Certification Authorities (ACACA 1995) produced the *Guidelines for Assessment Quality and Equity*. The primary focus of this document is to ensure quality, fair assessment in high stakes senior secondary assessment. This document includes 20 guidelines that address the quality of assessment methods, materials, and results.

Over a decade later, in the UK, the Assessment Reform Group (2008) issued *Changing Assessment Practices: Process, Principles and Standards*. The purpose of this document is to guide and support change in assessment practice in educational contexts. Four broad standards and associated guidelines address: classroom teachers, school management teams, national/local governing bodies, and policy makers. Within each standard, assessment is discussed generally and in terms of formative and summative uses.

In 2012, the Association for Educational Assessment-Europe (AEA-Europe) produced the *European Framework of Standards for Educational Assessment 1.0*. This framework is designated as a tool with seven core elements intended to support users of assessments as well as providers of assessment education and training. The framework incorporates both summative and formative uses of various types of assessments: standardized tests, classroom assessments, performance assessments, and assessments of learning outcomes of a program/curriculum.

3.3 Teacher accreditation- and certification-based assessment standards

In 2001, the National Board for Professional Teaching Standards (NBPTS) in the USA issued *What Teachers Should Know and Be Able to Do* to inform the National Board certification of American teachers. This document outlines five core propositions that illustrate the level of knowledge, skills, abilities, and commitments essential to quality teaching. Embedded within these propositions are statements pertaining to the assessment competencies demonstrated by effective teachers.

Seven years later, the National Council for the Accreditation of Teacher Education published the *Professional Standards for the Accreditation of Teacher Preparation Institutions* (NCATE 2008). NCATE is concerned with accountability and improvement in American teacher education programs. Within these standards, the assessment competencies necessary for pre-service teacher candidates are delineated.

Shortly after, the Interstate Teacher Assessment and Support Consortium (InTASC) (InTASC) (2011) released the *InTASC Model Core Teaching Standards: A Resource for State Dialogue*. This document is aimed at individuals and organizations responsible for the preparation, licensure, support, evaluation, and/or remuneration of teachers; its

goal is to articulate what effective teaching and learning look like. The InTASC standards are aligned with prominent national and state standards documents including the NBPTS teaching standards and the NCATE teacher accreditation standards.

On the other end of the globe, New Zealand (2008) issued *Graduating Teacher Standards* to guide the certification of new teacher entering the field. This document consists of seven standards with associated guidelines divided into three categories: Professional Knowledge, Professional Practice, and Professional Values and Relationships. Standards specific to assessment are embedded within these standards.

In 2012, the Department of Education in Australia published the *Australian Professional Standards for Teachers*. This document comprises seven standards with associated guidelines divided into three domains: Professional Knowledge, Professional Practice, and Professional Engagement. Assessment standards are incorporated into the guidelines. This standards document is used for accreditation of teacher education programs, licensure of new teachers, renewal of teacher registration, and recognition of exemplary teachers.

Finally, in 2012, the Department of Education in the UK published revised *Teacher Standards* to guide quality teaching and foster student achievement. These standards outline the minimum standards required for teacher certification. Within the standards document, teachers' assessment competencies are addressed.

3.4 Thematic analysis of standards

Eight themes were identified across the 15 assessment standards documents (see Table 2). These themes were (a) Assessment Purposes, (b) Assessment Processes, (c) Communication of Assessment Results, (d) Assessment Fairness, (e) Assessment Ethics, (f) Measurement Theory, (g) Assessment for Learning, and (h) Assessment Education and Support for Teachers. *Assessment Purposes* refers to choosing the appropriate form of assessment based on clearly stated instructional goals. *Assessment Processes* encompasses constructing, administering, and scoring assessment and interpreting assessment results to facilitate instructional decision-making. *Communication of Assessment Results* entails communicating assessment purposes, processes, and results to stakeholders. *Assessment Fairness* involves cultivating fair assessment conditions for all learners with sensitivity to student diversity and exceptional learners. *Assessment Ethics* means disclosing accurate information about assessments and protecting the rights and privacy of students that are assessed. *Measurement Theory* focuses on understanding psychometric properties of assessments (e.g., reliability and validity). *Assessment for Learning* describes the use of formative assessment during instruction to guide teacher practice and student learning. *Assessment Education and Support for Teachers* represents supporting teachers' assessment competency through explicit education opportunities or resources. These themes were analyzed across standards documents in terms of (a) shifts over time from 1990 to present and (b) variations across regions.

Shifts in assessment standards over time Our thematic analysis of all assessment standards, from 1990 to the present, revealed significant changes in the characterization of assessment competencies for teachers. We present our results in relation to the following temporal periods: 1990–1999, 2000–2009, and 2010–present (see Tables 3 and 4).

1990–1999 The primary themes identified in assessment standards documents from 1990 to 1999 were Assessment Purposes, Assessment Processes, Communication of Assessment Results, and Assessment Fairness (collectively 100 % of AFT et al. 1990; 100 % of JAC 1993; 80 % of ACACA 1995; 41.8 % of NCME 1995; and 75 % of AERA 1999). These early documents focused on helping teachers select and use assessments, primarily summative and standardized assessments, in order to make and communicate fair educational decisions about students. Hence, in the 1990s, assessment literacy focused heavily on summative forms of measurement with an emphasis on developing teachers' psychometric understandings. This focus is not surprising given its correspondence with the onset of the accountability movement across many parts of North America. Assessment standards that emphasized large-scale, standardized assessment further incorporated themes of Measurement Theory (5 % of ACACA 1995; 2.4 % of NCME 1995; and 25 % of AERA 1999) and Assessment Ethics (55 % of NCME 1995), including reliability, validity, norms, disclosure of information, and protecting students' rights and privacy. Only one document identified Assessment Education and Support for Teachers (2.3 % of NCME 1995), suggesting that while there were expectations for teachers' technical understandings about assessment, few standards were in place to support teacher learning in assessment.

2000–2009 In the following decade, Assessment Purposes, Assessment Processes, Communication of Assessment Results, and Assessment Fairness remained central themes in assessment standards documents for teachers; however, prominent new themes also emerged. Most notably, Assessment for Learning was identified in all three documents reviewed from this period (8.8 % of NBPTS 2001; 40.4 % of ARG 2008; and 7.1 % of NCATE 2008). The theme of Assessment for Learning (a) highlighted the importance of teachers' competency with practices including formative and diagnostic assessment, self- and peer-assessment among students, and teacher feedback to students and (b) extended the conception of assessment beyond summative and standardized uses. Assessment Education and Support for Teachers also emerged as a theme relevant to assessment competency (7.1 % of NCATE 2008), suggesting a growing awareness that teachers' competency in assessment needed to be coupled with formal provisions for teacher learning in assessment. Providing opportunities for teachers to cultivate assessment competency was recognized as a critical component in developing effective teachers. This finding makes sense given the emerging emphasis on Assessment for Learning during this period, as there was a greater emphasis on the integration of assessment with pedagogy and the use of assessment data to guide daily teaching and learning.

2010–present In recent years, assessment standards and integrated assessment standards reflect primarily an emphasis on Assessment for Learning. Assessment Purposes, Assessment Processes, Communication of Assessment Results, and Assessment Fairness remain critical aspects of assessment competency; however, Assessment for Learning has become a more dominant theme in modern assessment standards (27.4 % of Brookhart 2011; 17.6 % of JCSEE 2015; 7.9 % of Department of Education-UK 2012). Assessment Education and Support for Teachers has also increased as a critical component of teachers' assessment competency in recent documents (18.2 % of

Brookhart 2011; 5.9 % of JCSEE 2015). The *European Framework of Standards for Educational Assessment 1.0* (AEA-Europe 2012) is a notable exception; it is the only document since 2000 that does not include themes of Assessment for Learning or Assessment Education and Support for Teachers. The thematic nature of the European Framework is more congruent with the standards documents issued in the 1990s, with an emphasis on the selection and use of assessments to make educational decisions about students.

Variations across regions Looking across the five English-speaking regions (i.e., Australia, Canada, New Zealand, UK, and USA), current assessment standards (2000-present) are strikingly similar, with an emphasis on Assessment for Learning and the use of assessment data to guide daily classroom instruction and learning. There are, however, differences in the onset of this dominant theme across the countries examined in this research. Although the Assessment for Learning emerged strongly in the *Changing Assessment Practices: Process, Principles and Standards* document issued in the UK in 2008 (41.3 % of ARG 2008); this trend has only recently been mirrored in recent US assessment standards documents (27.4 % of Brookhart 2011; 17.6 % of JCSEE 2015). It appears that the work of the Assessment Reform Group, with its emphasis on formative assessment and assessment for learning (ARG 2008), informed the development of recent US assessment standards documents. In contrast, the *European Framework of Standards for Educational Assessment 1.0* (AEA-Europe 2012) has not fully followed the AFL trend.

Within teacher accreditation and certification-based standards documents, New Zealand's assessment standards for teachers (2008) mentioned Assessment for Learning once (3.4 %); while the USA led the way in assessment standards for teachers that emphasized the theme of Assessment for Learning (8.8 % of NBPTS 2001; 7.1 % of NCATE 2008; 8.6 % of InTASC 2011). These American documents highlighted the importance of preparing, certifying, and supporting teacher competency with respect to Assessment for Learning. Australia and the UK followed this trend by emphasizing Assessment for Learning in their most recent teaching standards documents (8.1 % of Department of Education-Australia 2012; 7.9 % of Department of Education-UK 2012).

3.5 Summary of assessment standards

Overall, results from our thematic analysis indicated that early documents (1990–1999) emphasized the selection and use of use assessments, primarily summative and standardized, in order to make and communicate fair educational decisions about students. Assessment for Learning emerged as a dominant theme in documents released after 2000, which coincides with the development of the Assessment Reform Group in the UK who emphasized the value of AFL. For example, the USA has since incorporated Assessment for Learning into teaching standards for preparation, certification, and ongoing education (NBPTS 2001; NCATE 2008; InTASC 2011). Finally, Assessment Education and Support for Teachers has also emerged as a new theme in documents released after 2000. These results highlight the fact that modern conceptions of assessment literacy entail articulations of supports for teacher learning that involve

coupling assessment for learning theory and practice with previously articulated summative-based assessment standards.

3.6 Assessment literacy: how do we measure it?

The second purpose of this study was to examine how existing instruments that aim to measure assessment literacy align with contemporary standards for teacher assessment literacy. In total, we reviewed eight instruments published post-1990 using a two-part analysis structure. First, we analyzed descriptive information about each instrument's design, guiding framework, and psychometric properties (see Table 5). We then analyzed the eight instruments in relation to the identified themes representing contemporary assessment literacy standards.

We, first, present descriptive information on the eight instruments in two categories: instruments that use the 1990 Standards (AFT et al. 1990) as their guiding framework and instruments based on other guiding frameworks.

Instruments based on the 1990 Standards The 1990 Standards for Teacher Competence in the Educational Assessment of Students provided the guiding framework for six of the instruments: the Assessment Literacy Inventory (ALI); the Assessment Practices Inventory (API); the Assessment in Vocational Classroom Questionnaire, Part II; the Classroom Assessment Literacy Inventory (CALI); the revised ALI; and the Teacher Assessment Literacy Questionnaire (TALQ). Of these instruments, the TALQ (Plake et al. 1993) served as the basis for three other instruments, namely the ALI, the CALI, and the revised ALI.

The TALQ is a 35-item, content based, instrument developed to measure in-service teachers' competency in the seven standards articulated in the 1990 Standards for Teacher Competence in the Educational Assessment of Students. The items were developed such that five items were developed for each standard. Initial psychometric properties for the instrument were determined from a sample of 555 in-service teachers. The internal consistency reliability estimate for this sample was found to be 0.54 and on average respondents received a score of 23.2 (SD=3.3) (Plake et al. 1993). In later years, the TALQ was administered to pre-service teachers (Campbell et al. 2002). However, for this administration, the TALQ was renamed the ALI. From a sample of 220 pre-service teachers, the internal consistency reliability estimate for this sample was found to be 0.74 and on average respondents received a score of 21, a slightly lower average than the previous study with in-service teachers (Campbell et al. 2002).

The CALI (Mertler 2003) was developed for use with both in-service and pre-service teachers. Once again, the TALQ served as the basis for this instrument. The CALI consisted of the same 35 content-based items, "with a limited amount of rewording (e.g., changing some names of fictitious teachers, changing word choice to improve clarity, etc.)" (Mertler 2003, p. 14). The instrument was administered to both in-service and pre-service teachers. From a sample of 197 in-service teachers, the internal consistency reliability estimate for this sample was found to be 0.57 and on average in-service respondents received a score of 22 (SD=3.4); whereas from a sample of 220 pre-service teachers, the internal consistency reliability estimate was found to be 0.74 and on average pre-service respondents received a score of 19 (SD=4.7) (Mertler 2003).

The revised ALI (Mertler and Campbell 2005) was developed in response to calls to revise the TALQ (e.g., Mertler 2003). As stated on the cover of this revised 35-item instrument, the instrument “consists of five scenarios, each followed by seven questions. The items are related to the seven ‘Standards for Teacher Competence in the Educational Assessment of Students.’ Some of the items are intended to measure general concepts related to testing and assessment, including the use of assessment activities for assigning student grades and communicating the results of assessments to students and parents; other items are related to knowledge of standardized testing, and the remaining items are related to classroom assessment” (Mertler and Campbell 2005, p. 26). As noted, although the 35 items were distributed among five scenarios, the allocation of five items per standard was also retained. The revised ALI was administered to 250 pre-service teachers. Within this sample, the internal consistency reliability estimate was found to be 0.74 and on average respondents received a score of 24 (SD=4.6) (Mertler and Campbell 2005).

The two remaining instruments that were developed using the 1990 Standards, the API and the Assessment in Vocational Classroom Questionnaire, Part II, were developed using Likert-type items. The API (Zhang and Burry-stock 1997) was developed to measure in-service teachers’ perceptions of their assessment skills. Each of the 67 items in the instrument used a 7-point scale that ranged from 1=not confident to 7=very confident. The API was administered to 297 in-service teachers. Using principal axis factoring with varimax rotation on this sample, items were grouped into seven subscales. The seven subscales were named: Perceived Skillfulness in Using Paper-Pencil Tests (16 items); Perceived Skillfulness in Standardized Testing, Test Revision, and Instructional Improvement (14 items); Perceived Skillfulness in Using Performance Assessment (10 items); Perceived Skillfulness in Communicating Assessment Results (9 items); Perceived Skillfulness in Nonachievement-Based Grading (6 items); Perceived Skillfulness in Grading and Test Validity (10 items); and Perceived Skillfulness in Addressing Ethical Concerns (2 items). Furthermore, with this sample, the internal consistency reliability estimates for subscales were found to range from 0.79 to 0.93, and the internal consistency reliability estimate for the entire instrument was found to be 0.97 (Zhang and Burry-stock 1997).

Finally, the Assessment in Vocational Classroom Questionnaire, Part II (Kershaw IV 1993) was developed to measure in-service teachers’ perceived level of competence in assessment activities. The 26 items in this instrument used a 5-point scale with the following descriptors: 1=not competent, 2=slightly competent, 3=moderately competent, 4=very competent, and 5=extremely competent. The items covered choosing assessment methods; developing assessment methods; administering, scoring and interpreting results; using assessment results; developing grading procedures; communicating assessment results; and identifying ethical issues. The maximum number of points a respondent could get with this instrument was 130 (i.e., 26 items × 5 points per item). When administered to 393 in-service teachers, the internal consistency reliability estimate was found to be 0.91, and the mean total score was found to be 97.0 (SD=12.9).

Instruments based on other frameworks The Assessment Self-Confidence Survey (Jarr 2012) was developed to measure teachers’ self-confidence with assessment-related practices. The instrument was developed following Bandura’s (2006) guidelines for

constructing self-efficacy scales. The instrument consists of 15 Likert-type items that use a 7-point scale (1=not confident at all to 7=very confident). The items covered interpreting standardized test results, using assessment results (both formative and summative classroom assessments), communicating assessment results, and adhering to legal and ethical obligations. The maximum number of points a respondent could get with this instrument was 105 (i.e., 15 items \times 7 points per item). When the instrument was administered to 201 in-service teachers, the internal consistency reliability estimate was found to be 0.90, and the mean total score was found to be 64.9 (SD=14.2).

The Measurement Literacy Questionnaire (Daniel and King 1998) was developed using assessment literature (e.g., Gullickson 1984; Kubiszyn and Borich 1996; Popham 1995). The 30 items were used to assess in-service teachers' assessment literacy, which was referred to as testing and measurement literacy in the study. Items covered assessment knowledge (e.g., what is a standardized test and what is the purpose of achievement tests), interpreting test results (e.g., percentiles, stanines, and test statistics), and communicating assessment results (e.g., use of the terms reliability and content validity). When the instrument was administered to 95 in-service teachers, the internal consistency reliability estimate was found to be 0.60, and the mean total score was found to be 18.2 (SD=3.3).

Overall, amongst the eight identified assessment literacy instruments, the instruments developed using the 1990 Standards for Teacher Competence in the Educational Assessment of Students (AFT et al. 1990) as a guiding framework continue to be most often cited and used. For example, the Assessment Practices Inventory has used in publications and dissertations (e.g., Braney 2010; Zhang and Burry-stock 2003). It is unsurprising that the majority of instruments are based on the 1990 standards as few recent instruments (i.e., post-2000) have been published and considered in this analysis. However, even newer instruments use the 1990 standards. Brookhart (2011) recognized the continued use of the 1990 Standards a guiding framework is particularly problematic given their lack of emphasis on formative assessment practices and standards-based education structures. In our subsequent analysis, we further analyze the eight assessment literacy measures to deduce the extent to which they address contemporary themes associated with assessment literacy.

3.7 Analysis of instruments by assessment literacy themes

We deductively analyzed the eight instruments in relation to the eight identified themes representing contemporary assessment standards. The frequency of each theme, in relation to the total number of items per instrument, was calculated and expressed as a percentage (Table 6). Across the eight instruments, the Assessment Processes theme was most commonly represented with frequencies ranging from 46.7 to 61.5 %. Other common themes, when represented, included the Communication of Assessment Results (range, 11.5 to 26.7 %), Assessment Ethics (range, 4.5 to 14.3 %), and Assessment Purposes (range, 3.3 to 14.3 %).

Given that the TALQ served as the basis for three other instruments (ALI, CALI, and revised ALI), four of the eight instruments (i.e., ALI, CALI, revised ALI, and TALQ) had identical frequency distributions, representing four contemporary themes: Assessment Processes (57 %), Assessment Purposes (14 %), Communication of

Assessment Results (14 %), and Assessment Ethics (14 %). The API and the Assessment Self-Confidence Survey were the only instruments that had items representing Assessment Fairness and Assessment for Learning themes; while only the Measurement Literacy Questionnaire had a high percentage of items (43.3 %) representing the Measurement Theory theme. Finally, the 67 API items represented seven out of the eight contemporary themes. Across all eight instruments, no items represented the Assessment Education and Support for Teachers theme.

3.8 Summary of assessment literacy instruments

Overall, results of our analysis indicate that the 1990 Standards for Teacher Competence in the Educational Assessment of Students (AFT et al. 1990) continues to be the predominant guiding framework used to create assessment literacy instruments. In relation to the representation of contemporary assessment standards within identified instruments, the Assessment Processes theme was found to be most commonly represented. When present within instruments, Communication of Assessment Results, Assessment Ethics, and Assessment Purposes themes were often equally represented. On the other hand, the Measurement Theory theme was only prominent in one instrument. Finally, while only two instruments represented Assessment Fairness and Assessment for Learning themes, no instruments represented the Assessment Education and Support for Teachers theme.

4 Discussion

Measuring and supporting teachers' assessment literacy have been the focus of educational policy and research since the early 1990s (Gotch and French 2014; Plake et al. 1993; Popham 2013; Stiggins 2004). Since the *Standards for Teacher Competence in Educational Assessment of Students* (AFT et al. 1990), researchers have aimed to characterize the multiple dimensions of "assessment literacy" through various assessment standards and work toward sound measures that could analyze teachers' strengths and weaknesses in this critical aspect of their practice. Through this research, we have analyzed temporal and geographic trends in assessment literacy standards and related measures. Specifically, our analysis included 15 assessment standards from five English-speaking countries plus mainland Europe and eight widely used assessment instruments.

Results from this study show a gradual shift in conceptions of assessment literacy over time. Initial standards emphasized teachers' abilities to construct, administer, and use primarily summative forms of assessment. Since 2000, most standards have integrated, to greater and lesser extents, the concept of assessment for learning and assessment education. Interestingly, measures of assessment literacy have not necessarily responded to this shift. This finding is not surprising as the majority of existing assessment literacy measures continue to use the 1990 Standards (AFT et al. 1990) as their guiding framework. We agree with previous researchers that the persistent use of the 1990 Standards is problematic as they do not fully recognize the formative role of assessment within highly diverse (i.e., socio-cultural and economic) contexts of standards-based education (Brookhart 2011; Gotch and French 2014). As a result, many assessment literacy measures appear to over represent the theme of Assessment

Processes with a continued focus on classroom summative assessment and standardized testing.

While there were some variations in the frequency representation of assessment literacy themes across standards documents, assessment standards were fairly consistent in their composition across regions, with the exception of the European Framework of Standards for Educational Assessment, which continued to primarily emphasize Assessment Processes. Across regions and type of standards document (i.e., government, research, and teacher certification/accreditation), there appears a growing emphasis on themes of Assessment for Learning and Assessment Education and Supports for Teachers. While the first of these is not surprising given the surge of literature on assessment for learning since Black and Wiliam's (1998) synthesis on feedback and learning, subsequent work of the Assessment Reform Group (2002, 2008) in the UK, and additional research that demonstrates the value of assessment for learning on student achievement, motivation, and metacognitive development (Crooks 1988; Earl 2003; Kluger and DeNisi 1996; Natriello 1987; Wiliam 2007). However, the second of these themes provides an interesting finding in light of measuring assessment literacy. If the aim is to use results from assessment literacy measures to support teachers in developing their assessment literacy through responsive and targeted teacher education, then it would be useful if the measure itself addressed teachers' learning experiences and preferences for assessment education. Coupled with data on their strengths and weaknesses in assessment, this information would enable purposefully designed teacher education that responds not only to teachers' learning needs but also to their learning preferences. We believe that it is through such a tailored approach that researchers and teacher educators can begin to curtail the persistent low assessment literacy rates that pervade amongst teachers (DeLuca 2012; Galluzzo 2005; MacLellan 2004; Mertler 2003, 2009; Plake et al. 1993; Volante and Fazio 2007; Zhang and Burry-stock 1997).

Our fundamental aim in conducting this research was to provide a starting point for the development of future assessment literacy measures that can be used for constructing responsive assessment education structures. Hence, results from this research yield five important recommendations for developing reliable measures that allow valid interpretations of teachers' assessment literacy. These recommendations are:

1. Predicate assessment literacy measures on contemporary assessment standards to promote greater validity of results. Measures should reflect the complexity of the assessment literacy construct as delineated through the eight themes identified in this research and adapted to regional assessment policies and priorities. Utilizing contemporary standards as a guiding framework for constructing measures will promote construct validity (Messick 1989; Kane 2006) by responding to the multiple professional responsibilities associated with assessment literacy. Adapting measures to regional policies and priorities will further enable geographic validity (Messick 1989).
2. Consider assessment education and support for teachers when constructing assessment literacy measures. Specifically, gaining information on teachers' preferences, experiences, and perceived effectiveness of assessment education structures would enable the development of data-based, responsive teacher education. Coupling items related to assessment literacy and assessment education would provide an information-rich basis to inform teacher learning in assessment.

3. Specify the focal teacher population for the developed instrument. Existing measures have primarily been pilot tested with in-service teachers, yet are used widely to measure the assessment literacy of pre-service teachers. These two teacher populations may have differing learning needs in assessment and value different learning structures. Hence, in developing assessment literacy measures, there is a need to test item stability across these two populations. This recommendation is particularly important given the emerging emphasis in standards documents on assessment education and supports for teacher learning.
4. Continue to work toward enhancing the reliability of measures. As recognized by Gotch and French (2014), there is a persistent need to enhance the reliability of assessment literacy indicators. That said, we acknowledge that the reliability of practical assessment literacy measures is challenging given the multi-dimensionality of assessment literacy (i.e., eight inter-connected themes). Hence, we recognize that reliability may be a persistent challenge for assessment literacy measures that aim to reflect the dimensionality of this construct while attending to issues of instrument feasibility and administration (i.e., length, duration, scoring, and delivery).
5. Establish the value and validity of assessment literacy instruments based on (a) a close coupling with both assessment standards (i.e., assessment research and theory) and teachers' actual assessment practices (i.e., correspondence between what teachers say they do/know in assessment and how they actually assess in their classrooms) and (b) consequences to teacher learning and professional assessment education (i.e., does the instrument provoke positive learning consequences for teachers based on responsive teacher education?).

5 Limitations

While this research included assessment standards from six regions (i.e., Australia, Canada, mainland Europe, New Zealand, UK, and USA) and eight assessment instruments, there are limitations to the study's methodology. First, only texts from English-speaking countries (plus mainland Europe) were used, eliminating the perspectives towards assessment standards and instruments from non-English speaking regions or additional English-speaking nations. Second, only national or inter-state standards were selected, which do not account for more local (i.e., state or province) documents that guide assessment practices. Finally, the majority of instruments used in phase 2 of this study were developed during the period of 1990–2000; hence, it is unsurprising that many assessment literacy instruments rely on the 1990 Standards. Newer instruments, including those currently under development, may reflect more contemporary assessment standards. Overall, these limitations suggest that although this research has provided a starting point for continued research on teacher assessment competency, it may not have included all potential assessment standards or instruments. Future research should continue to map assessment instruments as they develop to contemporary assessment standards. Further, as an assessment community, we should continue to trace the evolution of assessment standards at international, national, and more local levels to determine temporal and geographic priorities.

6 Conclusion

It is clear that we need to support teachers in their professional responsibility to be assessment literate (Popham 2004, 2013). Establishing a strong understanding of what assessment literacy is and how we measure it is a necessary step in this process. Based on this study, we offer the above recommendation for constructing assessment measures that (a) reflect the multiple dimensions of assessment literacy as exhibited in contemporary standards and (b) are reliable for different teacher populations (i.e., pre- and in-service teachers). We see value in developing and establishing validity evidence for sound measures that accurately characterize teachers' strengths and weaknesses in assessment. These measures can then form the basis for responsive teacher education that works to enhance teachers' assessment literacy and ultimately improve classroom assessment practices.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Retrieved from <http://www.teststandards.org/>.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington: National Council on Measurement in Education.
- Assessment Reform Group. (2002). *Assessment for learning: research-based principles to guide classroom practice*. UK: University of Cambridge, Assessment Reform Group.
- Assessment Reform Group. (2008). *Changing assessment practices: Process, principles and standards*. Retrieved from <http://www.aaa.org.uk/content/uploads/2010/06/ARIA-Changing-Assessment-Practice-Pamphlet-Final.pdf>.
- Association for Educational Assessment-Europe. (2012). *European framework of standards for educational assessment 1.0*. Rome: Edizioni Nuova Cultura.
- Australian Curriculum, Assessment and Certification Authorities. (1995). *Guidelines for assessment quality and equity*. Retrieved from <http://acaca.bos.nsw.edu.au/files/pdf/guidelines.pdf>.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents, adolescence and education* (pp. 307–337). Greenwich: Information Age Publishing.
- 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*.
- Black, P., & Wiliam, D. (1998). *Inside the black box: raising standards through classroom assessment*. Granada Learning.
- Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9, 27–40.
- Braney, B.T. (2010). *An examination of fourth grade teachers' assessment literacy and its relationship to students' reading achievement*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses database. (UMI No. 3434570).
- Brookhart, S. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30, 3–12.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: applicability to pre-service teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.
- Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, 91(6), 331–344.

- DeLuca, C. (2012). Preparing teachers for the age of accountability: toward a framework for assessment education. *Teacher Education Yearbook XXI: A Special Issue of Action in Teacher Education*, 34(5/6), 576–591.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education: aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64(4), 356–372.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17, 419–438.
- Department for Education-United Kingdom. (2012). *Teachers' standards*. Retrieved from <http://www.education.gov.uk/schools/teachingandlearning/reviewofstandards/a00205581/teachers-standards-1-sep-2012>.
- Department of Education-Australia. (2012). *Australian professional standards for teachers*. Retrieved from <http://www.teacherstandards.aitsl.edu.au/OrganisationStandards/Organisation>.
- Earl, L. (2003). *Assessment as learning*. Thousand Oaks: Corwin.
- Galluzzo, G. R. (2005). Performance assessment and renewing teacher education. *Clearing House*, 78(4), 142–45.
- Gardner, J. (2006). Assessment for learning: a compelling conceptualization. In J. Gardner (Ed.), *Assessment and learning* (pp. 197–204). London: Sage.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33, 14–18.
- Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *The Journal of Educational Research*, 77(4), 244–248.
- National Council on Measurement in Education (1995). *Code of professional responsibilities in educational measurement*. Retrieved from: http://www.natd.org/Code_of_Professional_Responsibilities.html.
- International Symposium for Classroom Assessment (ISCA). (n.d.). *History 2001 - Assessment for learning: international perspectives*. Retrieved from: <https://a4larchive.wordpress.com/about-2/history-2001/>.
- Interstate Teacher Assessment and Support Consortium (InTASC). (2011). *InTASC model core teaching standards: a resource for state dialogue*. Washington: Council of Chief State School Officers. Retrieved from [http://www.ccsso.org/Resources/Publications/InTASC_Model_Core_Teaching_Standards_A_Resource_for_State_Dialogue_\(April_2011\).html](http://www.ccsso.org/Resources/Publications/InTASC_Model_Core_Teaching_Standards_A_Resource_for_State_Dialogue_(April_2011).html).
- Jarr, K.A. (2012). *Education practitioners' interpretation and use of assessment results*. University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/3317>.
- Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada*. Retrieved from http://www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf.
- Joint Committee for Standards on Educational Evaluation. (2015). *Classroom assessment standards: practices for PK-12 teachers*. Retrieved from http://www.teach.purdue.edu/pcc/DOCS/Minutes/12-15_Handouts/2013-0116/JCSSE_Assessment_Standards.pdf.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport: American Council on Education.
- Kershaw IV, I. (1993). *Ohio vocational education teachers' perceived use of student assessment information in educational decision-making*. Ohio State University.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback intervention on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Kubiszyn, T., & Borich, G. (1996). *Educational testing and measurement: classroom application and practice* (5th ed.). New York: HarperCollins.
- MacLellan, E. (2004). Initial knowledge states about assessment: novice teachers' conceptualizations. *Teaching and Teacher Education*, 20, 523–535.
- Mertler, C.A. (2003). Pre-service versus in-service teachers' assessment literacy: does classroom experience make a difference? In Annual meeting of the Mid-Western Educational Research Association, Columbus.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101–113. doi:10.1177/1365480209105575.
- Mertler, C.A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: development of the assessment literacy inventory. In Annual meeting of the American Educational Research Association, Montreal.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- National Board for Professional Teaching Standards. (2001). *What teachers should know and be able to do*. Retrieved from http://www.nbpts.org/sites/default/files/documents/certificates/what_teachers_should_know.pdf.

- National Council for the Accreditation of Teacher Education (2008). *Professional standards for the accreditation of teacher preparation institutions*. Retrieved from: <http://www.ncate.org/public/standards.asp>.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155–175.
- New Zealand Teachers Council. (2008). *Graduating teacher standards*. Retrieved from <http://www.teacherscouncil.co.nz/>.
- Patton, M. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand OaksCA: Sage Publications.
- Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: a national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–39. doi:10.1111/j.1745-3992.1993.tb00548.x.
- Popham, W. J. (1995). *Classroom assessment: What teachers need to know*. Boston: Allyn-Bacon.
- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62, 82–83.
- Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). Boston: Pearson.
- Stiggins, R. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan*, 83, 758–765.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86, 22–27.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: implications for teacher education reform and professional development. *Canadian Journal of Education*, 30, 749–770.
- Wiliam, D. (2007). Content *then* process: teacher learning communities in the service of formative assessment. In D. B. Reeves (Ed.), *Ahead of the curve: the power of assessment to transform teaching and learning* (pp. 183–204).
- Willis, J. (2010). Assessment for learning as a participatory pedagogy. *Assessment Matters*, 2, 65–84.
- Zhang, Z., & Burry-stock, J.A. (1997). Assessment practices inventory: a multivariate analysis of teachers' perceived assessment competency. In Annual meeting of the American Educational Research Association, Chicago.
- Zhang, Z., & Burry-stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342.