



U.S. teachers' conceptions of the purposes of assessment



Nicole Barnes^{*}, Helenrose Fives, Charity M. Dacey

Montclair State University, 1 Normal Ave., Montclair, NJ 07043, USA

H I G H L I G H T S

- Investigated 179 Northeastern U.S. practicing teachers' conceptions of assessment.
- Three factors: assessment as irrelevant, valid for accountability, informing teaching and learning.
- Cluster analysis indicated three unique belief profiles among these participants.
- Findings support perspectives that teachers hold multiple beliefs simultaneously.

A R T I C L E I N F O

Article history:

Received 9 February 2016
Received in revised form
5 January 2017
Accepted 17 February 2017
Available online 26 March 2017

Keywords:

Practicing teachers
Classroom assessment
Conceptions of assessment
Teacher's beliefs

A B S T R A C T

Teachers' conceptions about assessment influence their classroom assessment practices. In this investigation, we examined 179 K-12 teachers' conceptions of the purposes of assessment from a person-centered perspective. An exploratory factor analysis of teachers' responses to the Conceptions of Assessment Instrument yielded a three-factor model: assessment as valid for accountability, improves teaching and learning, and as irrelevant. Next, we used cluster analysis to identify belief profiles of teacher groups: Cluster-1: Moderate, Cluster-2: Irrelevant, Cluster-3: Teaching and Learning. Within and across cluster comparisons revealed significant differences indicating that these are distinct profiles: teachers can, and do, hold multiple beliefs about assessment simultaneously.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

In the United States (U.S.), assessment has received a great deal of attention from both the popular press (Ravitch, 2011) and education scholars (Darling-Hammond, 2014). Assessment at the national, local, and classroom level should serve multiple purposes; most salient, assessment should be seen and used by teachers to inform their instructional activities (Black & William, 2009). However, research and theory into teachers' beliefs has established that teachers' beliefs about their professional tasks (e.g., instruction, assessment) influence the ways that they engage in their work (e.g., Barnes, Fives, & Dacey, 2015; Fives & Buehl, 2012). In other words, how teachers conceive of assessment shapes if (and how) they use assessment results and frames their selection and development of assessment tools in their classrooms (Brown, 2008). Moreover, these conceptions may also influence how teachers engage

students in assessment activities. Thus, as assessment in U.S. schools takes center stage in education policy, understanding the beliefs teachers hold about the purpose of assessment should become more relevant for educational researchers, teacher educators, and school instructional leaders because teachers' conceptions affect their assessment practices (Brown, 2008). As such, the purpose of this investigation was to explore teachers' conceptions of the purposes of assessment.

2. Teachers' beliefs about assessment

There is a long history of research into teachers' beliefs to understand how these constructs serve to influence teacher learning, practice, and understandings of their context (e.g., Fives & Buehl, 2012; Pajares, 1992; Woolfolk Hoy, Davis, & Pape, 2006). Pajares (1992) highlighted that teachers hold beliefs or conceptions about particular things and teachers filter new information, frame problem spaces, and guide actions through those beliefs (Fives & Buehl, 2012; Fives & Buehl, 2014). In the context of assessment, teachers' beliefs about the purposes of assessment may influence how teachers process new information about assessment practices/

^{*} Corresponding author.

E-mail addresses: barnesn@mail.montclair.edu (N. Barnes), fivesh@mail.montclair.edu (H. Fives), daceyc@mail.montclair.edu (C.M. Dacey).

principles or help them frame situations where assessments are examined, developed, or selected. If one believes that assessments are used to hold teachers and schools accountable, then teachers can use this perspective to frame or limit their focus when they need to make decisions about assessments in their classrooms. Finally, beliefs about assessment may also serve as guides for teacher action in how they influence the decisions teachers make about other classroom practices such as instructional techniques and motivational strategies.

Within the literature on teachers' beliefs, a variety of terms and definitions have emerged to describe this construct. For instance, Thompson (1992) and Brown (2004, 2006) used the term conceptions rather than beliefs. Conceptions are considered to reflect "a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like" (Thompson, 1992, p. 130). The use of conceptions allows for knowledge and beliefs to be synthesized into a single construct, and this ameliorates some of the theoretical and operational challenges that emerge when scholars attempt to differentiate the two. In the field of assessment, Brown (2004, 2006) established a line of research using this construct to explore and measure teachers' beliefs about assessment.

2.1. Conceptions of assessment

In our recent review of the research on teachers' beliefs about assessment, we framed these beliefs on a continuum from an extreme pedagogical to an extreme accounting (accountability) conception (see Fig. 1, Barnes et al., 2015). Our review drew heavily on the work of Brown and his colleagues, who have explored teachers' conceptions of the purpose of assessment in multiple international and diverse cultural contexts, but not the U.S. (e.g., Brown & Harris, 2009; Brown, 2004, 2006). Brown (2004, 2006) developed and used the Conception of Assessment (COA-III) instrument to assess teachers' conceptions of the nature of assessment. Brown based the COA-III on the extant literature to describe the purposes of assessment, such that assessment is used to (1) promote or advance teaching and learning, (2) ensure that students are held accountable for their learning, and (3) ensure that teachers and schools are held accountable for students' learning (e.g., Heaton, 1975; Torrance & Pryor, 1998). Extending from these three conceptions for the purpose of assessment, Brown (2004, 2006) suggested that a fourth conception of assessment is prevalent among teachers, namely that assessment is irrelevant. Based on this framework, Brown developed the four-factor COA measure to assess teachers' conceptions of assessment. Below we describe these four conceptions within our proposed continuum of assessment for pedagogical to accounting purposes (Barnes et al., 2015).

At the pedagogical end of the continuum, we hold that teachers conceive of assessment as serving the purpose of informing instruction and improving student learning. The purpose of

informing instruction and improving student learning aligns with Brown's (2006) factor that reflects assessment as improving teaching and learning. On the COA-III, 12 items are associated with the factor of improving teaching and learning. These items tap into conceptions of assessment for the following purposes: improving teaching (e.g., "Assessment information modifies ongoing teaching of students"), improving learning (e.g., "Assessment feeds back to students their learning needs"), and describing abilities (e.g., "Assessment measures students' higher order thinking skills") (Brown, 2006, p. 168). In addition, a fourth conception of assessment as valid (e.g., "Assessment results can be depended on") is associated with this factor (Brown, 2006, p. 168). Thus, conceptions of assessment as informing teaching and learning are integrated with the conception that assessment provides valid and potentially useful information (Brown, 2004). Brown (2004) argued that for assessment information to inform teaching and learning, teachers must be able to trust that information. As such, conceptions of assessment as valid are a prerequisite for teachers to conceive of assessment as applicable to teaching and learning activities.

Toward the middle of our continuum we include the perspective that assessment is used to hold students accountable (Barnes et al., 2015). The three items on the COA that assess this conception of assessment describe assessment as "placing students into categories," "assigning grades," and determining "if students meet qualification standards" (Brown, 2006, p. 168). The degree to which teachers engage these conceptions to inform either instruction or determine grades (and therefore hold students accountable for their learning), likely varies by the teacher and the specific assessment tool/technique employed. Therefore, we argue that responses to these items might reflect a mixed conception of assessment as addressing both the pedagogical and accounting perspectives. As such, we placed it in the middle on our continuum.

At the accountability end of our continuum lie the conceptions that assessment is for holding schools and teachers accountable for student performance (Barnes et al., 2015). Specifically, the instrument includes three items that assess conceptions of assessment as providing information on school performance as an "indicator of school quality," and "a good way to evaluate a school" (p. 168). These items then provide a perspective that the purpose of assessment is to evaluate and make judgments about school and teacher quality and performance. That is, the purpose of assessment is to hold schools accountable rather than to inform teaching and learning, and as such it is placed on the accounting end of our continuum.

Brown (2004, 2006) includes a fourth conception of assessment (the contention that assessment is irrelevant to teaching and learning) which we placed outside of our continuum (Barnes et al., 2015). The assessment is irrelevant to teaching and learning factor is qualitatively different from the previous conceptions because if one considers assessment as irrelevant to teaching, then assessment cannot be used for either pedagogical or accounting purposes.

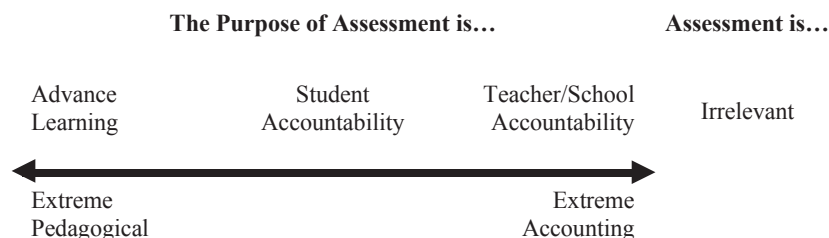


Fig. 1. Conceptions of assessment continuum.

Included in this conception are the perspectives that assessment is bad, ignored, and inaccurate. The instrument has nine items to assess this conception of assessment including assessment as bad for teachers, learners, and schools (e.g., “Assessment interferes with teaching,” p.168); something to be ignored or unused (e.g., “Assessment results are filed and ignored,” p.168); and inaccurate and consequently probably not useful for teaching and learning (e.g., “Assessment is an imprecise process,” p.168).

2.2. Research findings using COA

Following Brown's (2004, 2006) initial development of the COA-III, Brown and colleagues (e.g., Brown & Harris, 2009; Brown & Remesal, 2012) and other research teams (e.g., Segers & Tillema, 2011) have used the COA-III to examine teachers' conceptions of assessment in multiple contexts. In most cases when the COA-III was used researchers re-examined the factor structure of their data (e.g., Brown & Remesal, 2012) and in some cases developed new items and constructs to better fit the nature of the assessment context of their study (e.g., Brown, Hui, Yu, & Kennedy, 2011). Specifically, the COA-III (full and abridged versions) was administered (and as needed translated for participants) in New Zealand, Queensland (Australia), the Netherlands, Spain, China, Hong Kong, and India to investigate teachers' conceptions of assessment in countries with different cultural landscapes. The authors of these studies reported different numbers of emergent factors and variations in the assignment of items to similar factor titles.

Brown and Harris (2009) conducted a mixed methods investigation of Auckland, New Zealand teachers' conceptions of assessment. Teachers ($n = 161$) completed the COA-III. The quantitative findings indicated that Brown's (2009) four-factor structure for the COA-III was supported for use with these teachers. That is the same number of factors and item assignments as identified in the Brown (2006) study were replicated here. In contrast, when Segers and Tillema (2011) employed the COA-III with 351 secondary teachers in the Netherlands, they found a four-factor solution that differed from previous reports. They identified two factors that conceptually suggested that the purpose of assessment was to improve teaching and learning: (a) assessment informs performance and learning (11 items) and (b) assessment is of good quality (three items). They also identified one factor reflecting assessment as school accountability (three items) and one reflecting irrelevance, titled bad quality (three items). While these factors followed the same general trends and themes as those identified in the original measure, the assignment of items to factors differed. For example, the first factor, assessment informed performance and learning included items assigned by Brown (2004, 2006, 2009) to assessment improves education and student accountability. Similarly, the third factor, assessment is good quality included items from “improves education” (a sub-scale of improves teaching and describes abilities) and from “assessment is relevant” (a sub-scale of assessment is bad). In addition, none of the items from the assessment is irrelevant subscale were assigned to any factors in this investigation.

Differences in the number and make up of factors emerged in Brown and Remesal's (2012) investigation of preservice teachers from Spain ($n = 672$) and New Zealand ($n = 324$). Four different models were tested with the data from each of these sub-samples. Of note, the NZ sample indicated a five-factor solution and the Spanish sample indicated a four-factor solution similar to that found by Brown (2006) with two items dropped. After further analysis, Brown and Remesal (2012) elected to apply the five-factor solution to both sets of data for comparisons across countries. The five factors included: Assessment improves student learning and teaching (nine items from the assessment improves education subscales of describes abilities, improves learning, and improves

teaching, and one item from student accountability), assessment grades students (two student accountability items), assessment measures school quality validly (three school accountability items and three assessment is valid items), assessment is bad (three items, one from each of the assessment is irrelevant subscales) and assessment is ignored and inaccurate (six items including two from each of the assessment is irrelevant subscales). As with Segers and Tillema's (2011) study the items assigned to these factors differed from Brown's earlier models.

Brown and his colleagues (2011) compared the beliefs of primary and secondary teachers from Hong Kong ($n = 1014$) and Guangzhou, China ($n = 898$). However, before embarking on this investigation, awareness of the role of assessment in the cultural context of these regions led these researchers to re-conceptualize the COA-III and develop the Teachers' Conception of Assessment instrument for Chinese contexts (C-TCOA). Conceptually the new instrument included six constructs: the four COA-III constructs and two new constructs: assessment develops students into better people and assessment is used to control both students and teachers. In addition, new items and meaning were added to the existing COA-III factors. Analysis of data gathered using the C-TCOA revealed a seven-factor solution that was organized under three meta-factors: improvement, accountability, and irrelevance. Brown and colleagues' (2011) study illustrated the need for instruments to be sensitive to the cultural context in which the assessment is used and understood. More recently, Brown, Chaudhry, and Dhamija (2015) investigated 1,645 Northern Indian secondary (primarily private) school teachers' conceptions of assessment. Data were gathered using the COA-III plus the control items developed for the C-TCOA measure. Exploratory and confirmatory factor analysis resulted in a four-factor model, which suggested that teachers used assessment for improvement, viewed assessment as irrelevant, perceived assessment as a tool to control their lessons and teaching, and considered assessment as an indicator of school quality. The items assigned to these emergent factors varied somewhat from Brown's (2006) structure. Specifically, items assigned to the school quality factor included items previously associated with student accountability (one item) and improving education (two items from assessment is valid and one from assessment describes abilities).

The six studies described above illustrate the need to examine the factor structure each time data are collected to ascertain the underlying beliefs assessed. Although many of these studies were embedded in new contexts, there was an instance in which different factors and item assignments emerged across two studies with samples drawn from the same country (i.e., New Zealand), albeit exploring different populations (practicing: Brown & Harris, 2009 and preservice teachers: Brown & Remesal, 2012). Importantly, similar constructs seemed to emerge with each sample indicating that core conceptions of assessment were being assessed across these samples. However, these findings suggest that the nature of how these ideas exist and emerge when assessed must be examined with each new collection of data.

2.3. Variable versus person-centered approaches

Research in teachers' beliefs has long evidenced that teachers hold multiple, sometimes competing beliefs that influence their practice and interpretation of events (e.g., Fives & Buehl, 2012; Pajares, 1992). For instance, it is possible for a single teacher to view the purpose of assessment as both for improvement of student learning and for school accountability and as irrelevant (or any variation of these beliefs). Moreover, the configuration of these beliefs acting together or singly activated within the teacher may influence the instructional decisions he or she makes.

In fact, interview data from 26 teachers in Auckland, New Zealand revealed that teachers held conceptions of assessment reflecting all four factors, thus providing initial support for the multidimensional nature of teachers' conceptions of assessment (Brown & Harris, 2009). For example, teachers described assigning scores to elicit competition among students while also using assessment to give informational feedback on key skills. In another example, teachers reported that assessment provided evidence of school-level success and that assessment was used to hold schools accountable through external reporting. These qualitative findings illustrate the multidimensional nature of teachers' conceptions of assessment and suggest that teachers can (and do) have conflicting beliefs about assessment simultaneously.

With the exception of this qualitative study, the majority of investigators using the COA have employed a variable-centered approach to explore relations and differences with respect to conceptions of assessment across groups of teachers, and these studies have yielded valuable contributions to the extant literature (e.g., Brown & Harris, 2009; Brown, 2004). However, given the hypothesized multidimensional nature of teachers' beliefs, *person-centered* approaches are better able to capture the multiple perspectives (beliefs) that individuals may hold and allow for the analysis of belief profiles that emerge (e.g., Karabenick, 2003). Thus, in this study, we investigated teachers' conceptions of assessment using a person-centered approach guided by the following research questions:

- (1) What factor structure emerges when teachers respond to the COA-III?
- (2) What are teachers' conceptions of assessment and what, if any, relations emerge among those conceptions?
- (3) What, if any, profiles emerge with respect to these teachers' conceptions of assessment?

3. Method

3.1. Participants

Participants were 179 practicing K-12 teachers ranging in age from 20 to 70 years old. Participants taught primarily in the northeast region of the United States, with the largest group from New Jersey. They were 84% female, 88% White, and taught a range of grade levels and subject areas. Forty-four percent of participants had a master's degree with some additional graduate courses and 37% had been teaching between 11 and 20 years (See Table 1 for additional descriptive information).

3.2. The U.S. Context

In the U.S., free public school is available from kindergarten (approximately age 5) through 12th grade (approximately age 18), with some states and districts offering pre-kindergarten. Although there was no compulsory national assessment at the time of our data collection, the No Child Left Behind Act of 2001 required all public schools to administer statewide assessments, typically implemented with a single standardized test, annually to students in order to receive national funding. Individual states had autonomy to set learning goals/standards and to create their own standardized assessments. Continued state and federal funding was contingent on schools meeting "adequate yearly progress," which meant that students' standardized test results had to improve yearly. Failure to do so could result in loss of funding or state takeover of local schools. Student performance data on standardized tests were shared internally and externally with various stakeholders via student reports and the school's report card. Besides the annual standardized tests, however, most decisions

regarding assessment were made primarily at the local jurisdiction or school level. Teachers typically used classroom assessments formatively and summatively to gauge student learning and to improve their own instructional practices because the results from standardized assessments were not available until months after administration.

3.3. Conceptions of Assessment-III (COA-III)

We used the abridged (27 item) COA-III (Brown, 2006) to assess teachers' conceptions about the purpose of assessment. Items on the COA were developed to assess the theoretical conceptions of assessment as improvement (12 items), holding students accountable (three items), holding schools accountable (three items), and irrelevant (nine items). Items were assessed on a 6-point positively packed scale (1-strongly disagree, 2-mostly disagree, 3-slightly agree, 4-moderately agree, 5-mostly agree, and 6-strongly agree). "Positively packed" indicates that respondents have two negative (i.e., mostly and strongly disagree) and four positive (slightly, moderately, mostly, and strongly agree) options to select from for each statement. Confirmatory factor analyses of data conducted by the scale developer indicated that the four factors demonstrated good fit characteristics for primary teachers in both New Zealand ($\chi^2_{311} = 841.02$; RMSEA = 0.057; TLI = 0.87) and Queensland ($\chi^2_{311} = 1492.61$; RMSEA = 0.074; TLI = 0.80; Brown, 2006).

3.4. Procedures

We used a modified snowball sampling technique to recruit participants for this study. Members of our research team emailed and sent social media requests to teachers identified through personal contacts, professional online communities, and through our university-school partnership. We also asked anyone who responded to the survey to share the link to the survey with his or her colleagues. Interested teachers responded online to an anonymous questionnaire that resulted in 218 responses of which 206 responses were complete (i.e., no missing data). Upon reviewing the data, we determined that 27 participants were not K-12 teachers and their data were removed from the analysis. The present study was conducted by a team of teacher educators and their research team and was approved by our Institutional Review Board for research with Human Subjects.

4. Analyses and results

4.1. Exploratory factor analysis

The results of a confirmatory factor analysis indicated that the overall four-factor model found in Brown (2006) did not fit our data well ($\chi^2_{351} = 2785.81$; RMSEA = 0.109; TLI = 0.691). Therefore, we performed principal axis factoring (PAF) with promax rotation on participants' responses to the COA-III to identify a factor structure that best represented the data.

PAF is preferable to principal components analysis when the researcher's aim is to identify the underlying structure of the latent variables, while taking into account both the shared and error variance (Costello & Osborne, 2005). Promax is an oblique rotation, which allows the factors to correlate. Because our factors were moderately correlated, using an oblique rotation should "theoretically render a more accurate, and perhaps a more reproducible solution" (Costello & Osborne, 2005, p. 3). Furthermore, because the promax method retains a solution with the lowest kappa, it leads to factor structures that are "maximally distinguishable" (Matsunaga, 2010).

Table 1
Participant demographic descriptions.

| Variable | % | Variable | % |
|---|----|-------------------------------------|----|
| Age | | Gender | |
| 20–30 | 23 | Male | 16 |
| 31–40 | 31 | Female | 84 |
| 41–50 | 20 | Teaching Experience in Years | |
| 51–60 | 21 | 1–5 | 23 |
| 61–70 | 5 | 6–10 | 24 |
| Race/Ethnicity | | 11–20 | 37 |
| Caucasian | 88 | 20+ | 16 |
| African American | 6 | Teaching Context | |
| Hispanic | 5 | New Jersey | 61 |
| Asian | >1 | Maine | 5 |
| Native American | 1 | California | 4 |
| Other | >1 | New York | 4 |
| Education | | Maryland | 3 |
| Bachelor's Degree | 8 | Illinois | 3 |
| Bachelor's Degree + additional graduate courses | 18 | Pennsylvania | 2 |
| | | Massachusetts | 2 |
| Master's Degree | 28 | Colorado | 2 |
| Master's Degree + additional graduate courses | 44 | Washington | 2 |
| | | North Carolina | 1 |
| Doctoral Degree | 2 | Vermont | 1 |
| Grade Level | | Indiana | 1 |
| Early Childhood (PreK–2) | 11 | Georgia | 1 |
| Elementary (K–5) | 39 | Arizona | 1 |
| Middle Level (4–8) | 27 | Ohio | 1 |
| High School (9–12) | 24 | Iowa | <1 |
| Content Area | | Connecticut | <1 |
| More than one subject | 58 | Missouri | <1 |
| English | 13 | Michigan | <1 |
| Language Arts | 8 | New Hampshire | <1 |
| Mathematics | 6 | Utah | <1 |
| Science | 7 | Virginia | <1 |
| Social Studies | 5 | Louisiana | <1 |
| Other | 2 | Tennessee | <1 |

In order to identify the numbers of factors to extract we relied on Horn's (1965) parallel analysis. Thompson and Daniel (1996) argued that Horn's parallel analysis should be used in lieu of more traditional approaches such as the Kaiser-Guttman rule (i.e., eigenvalues greater than one). Horn's parallel analysis begins with principal axis factoring performed on randomly generated data sets. The eigenvalues of factors from these random data are then compared to the eigenvalues from the factors emerging from the study (i.e., actual) data. Factors in the study data are retained if the eigenvalue is greater than the eigenvalue from the random data. In this investigation, we compared the mean eigenvalues for 100 randomly generated data sets to the eigenvalues that emerged from the study data. Our parallel analysis indicated a three-factor solution. An examination of the scree plot further confirmed our decision to extract three factors and these factors collectively accounted for 51.54% of the variance. We examined the rotated factor matrix to determine which items to assign to each factor. Items with coefficients greater than |0.3| were assigned to that factor. In cases where items had more than one coefficient greater than |0.3|, we assigned that item to the factor upon which it had the highest loading. There was one exception to this in our analysis (described below).

We also reviewed Brown's (2006) descriptions of item assignments in relation to the emergent factor structure from our data. In column 2 of Table 2, we listed the factor and subfactor each item was assigned to by Brown (2006) and used this to guide our interpretation and naming of factors. We called Factor 1 *Assessment as valid for accountability*. Factor 1 included 10 items, six of these assessed beliefs about assessment as valid (e.g., Assessment results are trustworthy) and beliefs that describe what assessment does

(e.g., Assessment measures students' higher order thinking skills). In addition, three items considered by Brown (2006) to be school accountability (e.g., Assessment is a good way to evaluate a school) and one identified as student accountability (e.g., Assessment determines if students meet qualifications standards) items were assigned to this factor.

Interestingly, our factor structure suggested that two of Brown's (2006) *assessment describes abilities* subfactors could be interpreted as accountability items by the teachers in our sample (e.g., "Assessment is a way to determine how much students have learned from teaching" and "Assessment establishes what students have learned" (p. 168)). In the U.S., the national educational policy of No Child Left Behind (2001) has led to increased assessment of students, which is then used to evaluate teachers. Thus, the concept of assessment serving to measure what is learned from teaching or to establish what students have learned could be understood by the teachers in our sample to be items assessing beliefs about assessment as holding teachers accountable for student learning.

Factor 2, *Assessment improves teaching and learning*, included six items from Brown's (2006) *Assessment improves education* factor that assessed conceptions of assessment as improving or fostering teaching (e.g., Assessment is integrated with teaching practice) and learning (e.g., Assessment is integrated with teaching practice). Finally, the item "Assessment has little impact on teaching" had a coefficient of -0.31 for Factor 2 (Brown, 2006, p. 168). Conceptually, low scores on this item could be associated with high scores on the other items on Factor 2, however, reverse coding of COA-III items lack previous support. In addition, when we calculated the reliability statistic, Cronbach's alpha, for Factor 2 with and without this item, the reliability improved from 0.77 (with this item retained) to

Table 2
Pattern (structure) coefficients from the principal axis factoring with oblique rotation of the COA-III.

| Item | Brown (2006) Factor (Subfactor) | Factor | | |
|---|--|-------------------|-------------------|-------------------|
| | | 1 | 2 | 3 |
| Assessment is a good way to evaluate a school | Makes Schools Accountable | 0.91(0.82) | -0.04(0.43) | 0.16(-0.23) |
| Assessment is an accurate indicator of a school's quality | Makes Schools Accountable | 0.89(0.78) | -0.14(0.35) | 0.07(-0.27) |
| Assessment results are trustworthy | Improves Education (is Valid) | 0.72(0.80) | 0.04(0.50) | -0.15(-0.47) |
| Assessment provides information on how well schools are doing | Makes Schools Accountable | 0.68(0.60) | -0.02(0.31) | 0.17(-0.12) |
| Assessment is a way to determine how much students have learned from teaching | Improves Education (Describes Abilities) | 0.63(0.40) | 0.17(0.51) | 0.08(-0.26) |
| Assessment results are consistent | Improves Education (is Valid) | 0.63(0.70) | 0.07(0.46) | -0.06(-0.36) |
| Assessment determines if students meet qualifications standards | Makes Students Accountable | 0.62(0.68) | 0.18(0.50) | 0.12(-0.23) |
| Assessment results can be depended on | Improves Education (is Valid) | 0.57(0.73) | 0.20(0.57) | -0.10(-0.42) |
| Assessment establishes what students have learned | Improves Education (Describes Abilities) | 0.54(0.68) | 0.34(0.59) | 0.09(-0.25) |
| Assessment measures students' higher order thinking skills | Improves Education (Describes Abilities) | 0.48(0.62) | 0.28(0.54) | 0.05(-0.26) |
| Assessment information modifies ongoing teaching of students | Improves Education (Improves Teaching) | 0.02(0.46) | 0.73(0.76) | -0.04(-0.31) |
| Assessment feeds back to students their learning needs | Improves Education (Improves Learning) | 0.22(0.52) | 0.61(0.70) | 0.11(-0.21) |
| Assessment is integrated with teaching practice | Improves Education (Improves Teaching) | 0.15(0.51) | 0.56(0.68) | -0.07(-0.34) |
| Assessment provides feedback to students about their performance | Improves Education (Improves Learning) | 0.31(0.56) | 0.55(0.67) | 0.12(-0.21) |
| Assessment helps students improve their learning | Improves Education (Improves Learning) | 0.34(0.66) | 0.51(0.73) | -0.05(-0.38) |
| Assessment allows different students to get different instruction | Improves Education (Improves Teaching) | 0.16(0.46) | 0.49(0.60) | -0.04(-0.29) |
| Teachers conduct assessments but make little use of the results | Is Irrelevant (Ignored) | 0.27(-0.12) | -0.16(-0.25) | 0.68(0.62) |
| Assessment results are filed and ignored | Is Irrelevant (Ignored) | 0.24(-0.19) | -0.25(-0.35) | 0.68(0.66) |
| Assessment results should be treated cautiously because of measurement error | Is Irrelevant (Inaccurate) | -0.17(-0.21) | 0.34(0.04) | 0.57(0.52) |
| Assessment interferes with teaching | Is Irrelevant (Bad) | -0.22(-0.55) | -0.18(-0.50) | 0.52(0.68) |
| Teachers should take into account the error and imprecision in all assessment | Is Irrelevant (Inaccurate) | -0.25(-0.26) | 0.35(0.03) | 0.49(0.47) |
| Assessment is an imprecise process | Is Irrelevant (Inaccurate) | -0.41(-0.53) | 0.15(-0.26) | 0.48(0.61) |
| Assessment is unfair to students | Is Irrelevant (Bad) | -0.26(-0.58) | -0.23(-0.53) | 0.43(0.66) |
| Assessment forces teachers to teach in a way against their beliefs | Is Irrelevant (Bad) | -0.19(-0.46) | -0.15(-0.41) | 0.42(0.56) |
| Assessment is assigning a grade or level to student work | Makes Students Accountable | 0.13(-0.04) | -0.04(-0.09) | 0.34(0.30) |
| Assessment places students into categories | Makes Students Accountable | 0.31(0.12) | -0.08(-0.03) | 0.33(0.29) |
| Assessment has little impact on teaching | Is Irrelevant (Ignored) | 0.24(-0.04) | -0.31(-0.26) | 0.26(0.25) |
| Eigenvalues | | 9.53 | 2.63 | 1.76 |
| Percentage of variance explained (%) | | 35.30 | 9.73 | 6.51 |

Extraction Method: Principal Axis Factoring.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Note. Items with bolded coefficients were used in the calculation of means, standard deviations, and reliability coefficients. COA III = Conceptions of Assessment (Brown, 2006).

a more acceptable 0.85 (when this item was excluded). Therefore, for conceptual and analytic purposes, we excluded this item from further analyses.

We named Factor 3, *Assessment as irrelevant*. Ten items were assigned to this factor. These included eight items Brown (2006) identified in his analysis as belonging to the irrelevant factor (i.e., all but one) and two student accountability items. The irrelevant items included statements that assessment is ignored (e.g., Assessment results are filed and ignored), inaccurate (e.g., Assessment is an imprecise process), and bad (e.g., Assessment interferes with teaching; Brown, 2006, p. 168). Collectively, these items suggest that assessment of this nature cannot be relied on to make decisions. In addition, two of the items Brown (2006) described as student accountability items were assigned to this factor: "Assessment is assigning a grade or level to student work" and "Assessment places students into categories" (p. 168). As such, for the teachers in our sample items assessing beliefs about assessment as assigning grades or categorizing students were associated with issues of irrelevance; namely that assessment can be ignored, it is inaccurate or imprecise, and is not helpful for teaching.

4.2. Descriptive analyses

Descriptive analyses (i.e., means, standard deviations, reliability estimates, and correlations) were calculated to determine teachers' conceptions of assessment (i.e., Research Question 2; see Table 3). Our teachers indicated that they slightly (score of 3) to moderately (score of 4) agreed with constructs of *Assessment as valid for accountability* ($M = 3.625$, $SD = 0.734$), *Assessment improves teaching and learning* ($M = 3.254$, $SD = 0.752$), and *Assessment is irrelevant* ($M = 3.526$, $SD = 0.982$). Cronbach's reliability statistics

for these three factors were acceptable (Factor 1, $\alpha = 0.91$; Factor 2, $\alpha = 0.85$; Factor 3; $\alpha = 0.80$).

Table 3 also provides bivariate correlations across the three emergent factors. These factors demonstrated statistically significant correlations at the $p \leq 0.01$ significance level. Specifically, a moderately positive association ($r = 0.72$) was noted between Factor 1 and Factor 2. Relatively small negative associations emerged between the conceptions *Assessment is irrelevant* (Factor 3) and *Assessment as valid for accountability* (Factor 1; $r = -0.44$) and *Assessment improves teaching and learning* (Factor 2, $r = -0.45$). Thus, the more one viewed assessment as irrelevant the less likely he or she reported that assessment was valid for accountability purposes or that assessment could be used to improve teaching and learning.

4.3. Cluster analysis

We used cluster analysis to determine emergent profiles of teachers with respect to their conceptions of assessment based on their factor scores (Research Question 3). Initial clusters were formed using Ward's minimum variance hierarchical clustering technique to recover the underlying structure of the data and minimize the within-cluster differences (Atlas & Overall, 1994; Ward, 1963). To identify the number of clusters we followed a two-step process. First, we reviewed the dendrogram for gaps among potential cluster groups, which indicated a three-, four-, or five-cluster solution (Olson & Biolsi, 1991). Next we calculated the Calinski-Harabasz criterion (i.e., the variance ratio criterion, VRC), which is an appropriate method for determining clustering solutions with squared Euclidean distances (Calinski & Harabasz, 1974). The three-cluster solution yielded the largest VRC index value (419.62), which meant it had the largest between-cluster variance

Table 3
Means (standard deviations), reliability coefficients, and correlations among the COA-III Emergent Factors.

| Factor | No. of Items | M(SD) | Alpha | 1 | 2 | 3 |
|---|--------------|------------|-------|---------|---------|------|
| 1-Assessment as Valid for Accountability | 10 | 3.15(0.84) | 0.91 | 1.00 | | |
| 2-Assessment Improves Teaching and Learning | 6 | 3.59(0.96) | 0.85 | 0.72** | 1.00 | |
| 3-Assessment is Irrelevant | 10 | 3.62(0.76) | 0.80 | -0.44** | -0.45** | 1.00 |

** $p \leq 0.01$.

compared to within-cluster variance and thus represented the optimal number of clusters (Calinski & Harabasz, 1974). We used discriminant function analysis to validate the three-factor cluster solution (Romensburg, 1984). The three-factor model successfully predicted cluster membership 74.3% of the time and offered the most theoretically meaningful and interpretable description of the unique profiles that emerged (Fig. 2).

We performed a one-way multivariate analysis of variance with cluster membership serving as the independent variable and the three factors as the dependent variables. A significant multivariate effect was identified [$F(6, 348) = 35.39; p = 0.000; \eta = 0.38$]. Examination of the univariate tests indicated significant differences between the groups relative to *Assessment as valid for accountability* [$F(2, 176) = 55.11; p < 0.001; \eta = 0.39$], *Assessment improves teaching and learning* [$F(2, 176) = 122.42; p < 0.001, \eta = 0.58$] and *Assessment is irrelevant* [$F(2, 176) = 26.22; p < 0.001, \eta = 0.23$]. Because Levene's test was nonsignificant for all three factors, we used Scheffe's Test to determine differences across clusters in the post hoc analyses. All clusters demonstrated significant differences on all three factors at the $p < 0.001$ level.

We also conducted within-cluster comparisons to inform our characterization of each cluster. We used paired t -tests to compare the conceptions of assessment within each cluster (i.e., three dependent paired t -tests per cluster). We used a Bonferroni adjustment and a more conservative p -value criterion (i.e., $p = 0.05/3$ to determine statistical significance and to control for Type I error). The small boxes at the base of each cluster in Fig. 2 illustrate the significant differences within clusters. Consideration of differences within and across clusters informed our descriptors for each cluster. In addition, we provide demographic summaries of our three clusters in Table 4.

We dubbed Cluster 1 *Moderate*. Teachers in this cluster reflected 51% ($n = 91$) of our overall sample and had a moderate conception of assessment. Relatively moderate scores on all three factors suggested that teachers in this cluster perceived assessment as serving a number of purposes none of which were overwhelmingly salient. An examination of within cluster differences suggested these teachers had significantly higher beliefs in *Assessment is irrelevant* than they did for *Assessment improves teaching and learning* ($t(90) = 4.85; p < 0.001$) or for *Assessment as valid for accountability* ($t(90) = 2.50; p = 0.014$). Descriptively, cluster 1 was 87% female, 46% had a master's degree plus additional coursework, and 45% had approximately 11–20 years teaching primarily middle and high school students.

Cluster 2 was described as *Irrelevant*. Teachers in this cluster ($n = 37, 21\%$ of our sample) indicated significantly higher scores for *Assessment is irrelevant* in comparison to the other clusters. Further, teachers in this cluster reported the lowest scores of *Assessment as valid for accountability* and *Assessment improves teaching and learning*. This finding suggested that teachers in this cluster had a strong tendency to question the accuracy of assessments and saw it as interfering with or irrelevant to their teaching practice. Within this factor, scores for *Assessment is irrelevant* conceptions were significantly higher than conceptions of *Assessment as valid for accountability* ($t(36) = 8.85; p < 0.001$) or *Assessment improves teaching and learning* ($t(36) = 10.22; p < 0.001$). There was no statistically significant difference between their perceptions of *Assessment as valid for accountability* or *Assessment as improving teaching and learning* ($t(36) = 1.10; p = 0.277$). Thus, they strongly perceived assessment to be irrelevant and had little confidence that assessment could be used for instructional improvement or school accountability. Descriptively, teachers in this cluster were

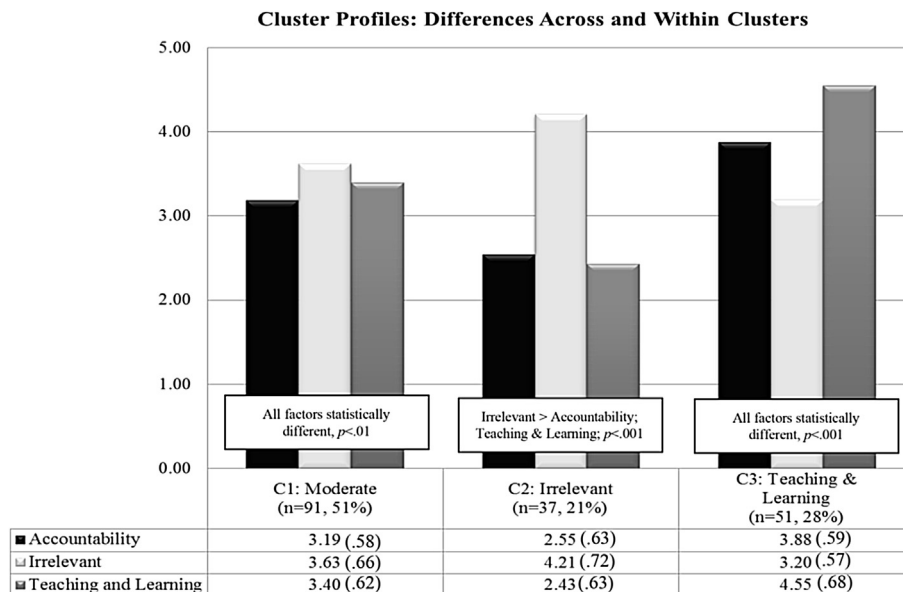


Fig. 2. Cluster profiles: differences across and within clusters.

Table 4
Descriptive statistics by cluster.

| Demographics | Clusters | | | | | |
|---------------------------------|--------------------------|-----------------|----------------------------|-----------------|-------------------------------------|-----------------|
| | C1: Moderate (n = 91) | | C2: Irrelevant (n = 37) | | C3: Teaching and Learning (n=51) | |
| | n | % of Cluster | n | % of Cluster | n | % of Cluster |
| Sex | | | | | | |
| Male | 12 | 13% | 7 | 19% | 15 | 29% |
| Female | 79 | 87% | 30 | 81% | 36 | 71% |
| Grade Level Teaching | | | | | | |
| Early Childhood (Pre-k-2) | 8 | 9% | 3 | 8% | 2 | 4% |
| Elementary (K-5) | 19 | 21% | 16 | 43% | 6 | 12% |
| Middle Level (4–8) | 32 | 35% | 10 | 27% | 19 | 37% |
| High School (9–12) | 32 | 35% | 8 | 22% | 24 | 47% |
| Years Teaching | | | | | | |
| 1-5 | 18 | 20% | 6 | 16% | 18 | 35% |
| 6-10 | 16 | 18% | 9 | 24% | 13 | 26% |
| 11-20 | 41 | 45% | 14 | 38% | 12 | 24% |
| 20+ | 16 | 18% | 8 | 22% | 8 | 16% |
| Education Completed | | | | | | |
| Bachelor's Degree | 10 | 11% | 4 | 11% | 5 | 10% |
| Bachelor's Degree + Add Courses | 17 | 19% | 4 | 11% | 9 | 18% |
| Master's Degree | 19 | 21% | 9 | 24% | 15 | 29% |
| Master's Degree + Add Courses | 41 | 45% | 20 | 54% | 21 | 41% |
| Doctoral Degree | 4 | 4% | 0 | 0% | 1 | 2% |

predominantly female (81%), 54% had completed a master's degree plus additional coursework, and 43% were elementary school teachers, the majority (38%) of these teachers had 11–20 years of experience.

We described Cluster 3 as *Teaching and learning*, which included 51 (28% of our sample) teachers who reported the perspective that the purpose of assessment was essentially to improve instruction and student learning. Within-cluster comparisons indicated that *Assessment improves teaching and learning* beliefs were significantly higher than beliefs for *Assessment as valid for accountability* ($t(50) = 8.38; p < 0.001$) and *Assessment is irrelevant* ($t(50) = 9.15; p < 0.001$). Teachers in this cluster were also more likely to believe in *Assessment as valid for accountability* than *Assessment is irrelevant* ($t(50) = 6.00; p < 0.001$), suggesting that this group was more likely to see different assessments as relatively useful. Descriptively cluster 3 was most similar to cluster 1; the majority of teachers were female (71%), most had completed coursework beyond a master's degree (41%), and 47% were teaching high school.

We used multinomial regression to test for cluster differences on the four dependent variables presented descriptively above and in Table 4 (i.e., gender, grade level taught, years teaching, and educational level completed) because our independent and dependent variables were both nominal (Johnson & Wichern, 2007). There were no significant differences between clusters on gender ($\chi^2 = 5.39; df = 2; p = 0.068$); number of years teaching ($\chi^2 = 10.08; df = 6; p = 0.122$); or level of education completed ($\chi^2 = 5.82; df = 8; p = 0.667$). Teachers did differ by cluster in terms of the grade level they taught ($\chi^2 = 14.91; df = 6; p = 0.021$) with teachers in cluster 2 (i.e., irrelevance) significantly more likely to teach elementary school than teachers in cluster 3 (i.e., teaching and learning).

5. Discussion and implications

In this investigation, we explored teachers' conceptions of assessment both to extend the existing work in this area, and to explore how these beliefs may configure from a person-centered perspective. We used exploratory principal axis factoring on the data we gathered from participants' responses to the COA-III to

ascertain the underlying structure of teachers' conceptions of assessment. Our analysis suggested a three-factor solution that, although different from prior work, maintains some conceptual relevance and is reflective of the conceptions of assessment continuum we described in our introduction (see Fig. 1). Factor 1, *Assessment as valid for accountability*, is reflective of the accounting end of our continuum while Factor 2, *Assessment improves teaching and learning*, reflects the pedagogical end of the continuum. Finally, Factor 3, *Assessment is irrelevant*, maps onto the same outlying concept in our continuum. Of note in our theoretical continuum we placed assessment as holding students accountable in the center of the figure, our empirical analyses revealed that items associated with this factor in Brown's (2004, 2006) work were split onto our emergent factors of *Assessment is valid for accountability* and *Assessment improves teaching and learning*. Thus, the items considered to tap assessment for student accountability may be seen by teachers as a middle-ground between assessment for teaching and learning and the more extreme accountability end of assessment that is present in many U.S. teaching contexts. Furthermore, our results echo prior research in which the researchers found different factor structures and factor loadings when the COA-III was administered to participants in varying contexts (e.g., Brown & Remesal, 2012; Segers & Tillema, 2011).

At the factor level, interesting configurations of items emerged that were distinct from those found in prior investigations using this measure. For instance, Factor 1, *Assessment as valid for accountability*, included items previously assigned to Brown's factors: *Assessment makes schools accountable* and *Assessment makes students accountable*, as well as items that described beliefs about the quality (validity) of assessment and items that described what assessment does (both previously assigned to *Assessment improves education*). It is noteworthy that our participants associated beliefs about assessment as accountability with issues of assessment validity, which possibly indicates that for assessment to serve accounting purposes necessitates that the assessment must yield valid inferences.

Assessment improves teaching and learning (Factor 2) included items related to improving teaching and learning. These items

seemed to emphasize formative practices that are distinct from accountability purposes, reflecting the extreme pedagogical end of our continuum. Moreover, there seemed to be a distinction in the level of the assessment, (school versus classroom level) with school level items associated with Factor 1 and classroom level items associated with Factor 2. It could be that the teachers in this sample associated issues of validity with standardized assessments used for accounting purposes. In contrast, when they considered classroom level assessments teachers may not have considered issues of assessment properties as readily. These findings contrast with Brown (2006) in which he noted that items related to assessment as valid were associated with *Assessment improves education*, suggesting that for an assessment result to be of use to a teacher or student it must be valid. The difference across these samples may be related to the cultural context of assessment at the national or district level. Alternatively, this may also be reflective of the quality of teacher preparation and continuing professional development as related to classroom assessment. In most U.S. teacher education programs, preservice teachers receive one course, at most, in assessment practices, or in many cases, this is one topic in another course. Thus, there are concerns about the depth of assessment literacy U.S. teachers may have, such that issues of validity and reliability are associated only with standardized assessments rather than classroom level measures used and frequently developed by teachers (Popham, 2011).

Items associated with the *Irrelevant* factor highlighted that assessment can be inaccurate, ignored, and lead to bad outcomes for students and teachers. Of note, items associated with the purpose of assessment as categorizing and grading students were also associated with this *Irrelevant* factor. It may be that the teachers responding to this instrument associated categorizing and grading students (*Assessment makes students accountable*) with poor assessment quality or as something that was detrimental to the goals of teaching and learning.

Another possible explanation for the overall variation in factor structure is methodological. The following prompt was used when administering the COA-III: “This survey asks about your beliefs and understandings about assessment, whatever that term means to you. Please answer the questions using your own understanding of assessment” (Brown, 2006). The nature of this prompt leaves us to question whether teachers are describing what they think assessment should be “in theory” rather than how they currently practice it in schools. In addition, “assessment” includes a myriad of objects and activities in the U.S. that include classroom, school, district, and state level practices. It is unclear which of these assessment activities teachers were considering in response to the items on the COA-III. Moreover, the way the items are worded (e.g., “assessment information modifies ongoing teaching of students” or “teachers conduct assessments but make little use of their results”) may lead teachers to respond based on their general conceptions of other teachers (or most) teachers’ assessment practices instead of their own personal practices. In fact, in prior research that had used both “I” and “Teachers” statements in the same instrument, researchers found that the way the item was worded led participants to adopt either a personal or teaching orientation when responding (Gibson & Dembo, 1984). Thus, future research using this tool in the U.S. should consider revising the instructions, providing a definition of what is meant by “assessment,” and using “I” statements.

Brown (2004) and Brown and colleagues (2015) have begun to address some of these issues in other contexts (i.e., New Zealand and Northern India respectively). In these investigations teachers were asked to either identify the type of assessment they were thinking about (Brown, 2004) or to consider a particular kind of

assessment (internally determined school based assessments versus externally mandated public examinations; Brown et al., 2015) when responding to the COA-III. As the high-stakes testing movement expands in the U.S., we feel that it may be necessary to clarify further in the use of the COA-III how teachers are conceiving of assessment. Brown’s (2004) and Brown et al., (2015) work serve as sound exemplars for how scholars may approach these challenges.

The cluster analysis results provide evidence for and highlight that teachers can, and do, hold multiple beliefs about assessment simultaneously. Teacher educators should consider the nature of these beliefs working in tandem with the teacher’s self-system¹ during professional learning experiences. Further, the patterns of beliefs suggest that for assessment-oriented interventions different instructional approaches may be needed for teachers holding varied perspectives. Teachers in Cluster 1 (i.e., Moderate) and Cluster 2 (i.e., Irrelevant) both reported significantly stronger agreement with items associated with the *Assessment as irrelevant* factor in comparison to the other two factors. Participants in these two clusters comprised 72% of our overall sample. Thus, the majority of our sample strongly supported beliefs that assessment is inaccurate, ignored, and bad. This finding suggests that many teachers see assessment as irrelevant/interfering with teaching and questionable in terms of quality. Therefore, teacher educators at the preservice or practicing teacher level may need to first combat issues of assessment quality or at the very least perceptions of assessment quality on the part of teachers.

Researchers need to tease out how these conceptions of assessment influence practice and identify potentially adaptive patterns of beliefs for positive instructional practice and openness to learning opportunities. Preservice teachers may hold naïve beliefs about assessment based on a limited understanding of the nature of assessment processes beyond their own experiences as students. Future research in the U.S. should engage in comparative studies of preservice and practicing teachers’ conceptions of assessment and breadth of knowledge regarding assessment practices. Further, meaningful learning trajectories for the development of sound assessment practices and the ability to critically review and make sense of assessment tools and results should also consider the ways that beliefs might change with respect to the purpose of assessment.

Finally, our work underscores that teachers hold multiple beliefs about assessment. In all of the clusters teachers indicated that they conceived of assessment as for improvement, accountability, and as irrelevant to their practice. Thus, when working with preservice and practicing teachers it is important to consider the configuration of their beliefs about assessment and to help them to make sense of these beliefs that may filter and guide their interpretation of assessment information.

6. Limitations

The sample was a non-random, convenience sample, in which teachers were largely from New Jersey, and therefore the results are not representative of or directly generalized to the larger U.S. or global teaching population. The nature of the data collection prevented us from gathering other forms of data (e.g., think aloud interviews) that would have allowed for triangulation of the findings or further investigations into teachers’ conceptions of

¹ Bandura defined self-system in the context of social learning theory as comprised of “cognitive structures and sub functions for perceiving, evaluating, and regulating behavior” (Bandura, 1978, p. 344). Here we refer to the self-system as teachers’ prior knowledge, beliefs, values, and experience of practice.

assessment. In addition, our recruitment procedures included appealing to professional listservs and social media pages, which prevented us from reporting response rate information. Thus, the sample of teachers who participated in this investigation may be unique in that they are willing to complete our questionnaire. However, the anonymity of the questionnaire may have had the advantage of garnering more realistic responses from participants.

7. Conclusion

Using exploratory principal axis factoring and cluster analysis, we examined teachers' conceptions of assessment based on their responses to the COA instrument. We found that teachers perceive assessment as valid for accountability, informing teaching and learning, and irrelevant. Further, we found that teachers hold multiple, sometimes competing beliefs about assessment, which may influence their engagement with varied assessment practices and willingness to engage in ongoing professional learning on this topic.

References

- Atlas, R. S., & Overall, J. E. (1994). Comparative evaluation of two superior stopping rules for hierarchical cluster analysis. *Psychometrika*, 59(4), 581–591. <http://dx.doi.org/10.1007/BF02294393>.
- Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, 33, 344–358.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives, & M. G. Gill (Eds.), *The handbook of research on teachers' beliefs* (pp. 229–233). New York, NY: Routledge.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <http://dx.doi.org/10.1007/s11092-008-9068-5>.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318. <http://dx.doi.org/10.1080/0969594042000304609>.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166–170. <http://dx.doi.org/10.2466/pr0.99.1.166-170>.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Brown, G. (2009). Teachers' self-reported assessment practices and conceptions: Using structural equation modelling to examine measurement and structural models. In T. Teo, & M. S. Khine (Eds.), *Structural equation modelling in educational research: Concepts and applications* (pp. 243–266). Rotterdam, The Netherlands: Sense Publishers.
- Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi experimental study of Indian teachers in private schools. *International Journal of Educational Research*, 71, 50–64.
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation*, 6(12), 68–91.
- Brown, G. T. L., Hui, S. K. F., Yu, 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. <http://dx.doi.org/10.1016/j.ijer.2011.10.003>.
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15(1), 75–89. http://dx.doi.org/10.5209/rev_SJOP.2012.v15.n1.37286.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27. DOI: <http://dx.doi.org/10.1080/03610927408827101>.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10(7). Retrieved from <http://pareonline.net/getvn.asp?v=10&n=7>.
- Darling-Hammond, L. (2014). *Next generation assessment: Moving beyond the bubble test to support 21st century learning*. Hoboken, NJ: John Wiley & Sons.
- Fives, H., & Buehl, M. (2012). Spring Cleaning for the “messy” construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Volume 2 individual differences and cultural and contextual factors* (pp. 471–499). Washington: American Psychological Association.
- Fives, H., & Buehl, M. M. (2014). Exploring differences in practicing teachers' valuing of pedagogical knowledge based on teaching ability beliefs. *Journal of Teacher Education*, 65(5), 435–448. <http://dx.doi.org/10.1177/0022487114541813>.
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569–582. DOI: <http://dx.doi.org/10.1037/0022-0663.76.4.569>.
- Heaton, J. B. (1975). *Writing English language tests*. London: Longman.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. doi: <http://dx.doi.org/10.1007/BF02289447>.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology*, 28(1), 37–58. [http://dx.doi.org/10.1016/S0361-476X\(02\)00012-7](http://dx.doi.org/10.1016/S0361-476X(02)00012-7).
- Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts and how-to's. *International Journal of Psychological Research*, 3(1), 97–110.
- Olson, J. R., & Biolsi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 240–285). Cambridge, England: Cambridge University Press.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332. <http://dx.doi.org/10.3102/00346543062003307>.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265–273. doi: <http://dx.doi.org/10.1080/08878730.2011.605048>.
- Ravitch, D. (2011). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Romensburg, H. C. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime Learning Publications.
- Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, 37, 49–54. <http://dx.doi.org/10.1016/j.stueduc.2011.03.008>.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). Reston, VA: National Council of Teachers of Mathematics.
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview of some guidelines. *Educational and Psychological Measurement*, 56(2), 197–208. DOI: <http://dx.doi.org/10.1177/0013164496056002001>.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Oxford: Taylor & Francis.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <http://dx.doi.org/10.1080/01621459.1963.10500845>.
- Woolfolk Hoy, A., Davis, H., & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander, & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 715–738). Mahwah, NJ: Lawrence Erlbaum.