



Exploring Filipino Teachers' Conceptions of Assessment

Christine Joy A. Ballada
De La Salle University, Manila

Marie Antoniette C. Aliño
St. Paul University-Quezon City

Abstract

Teachers' knowledge and beliefs about the teaching and learning process influence their classroom practices. One important component of this teaching-learning process that is affected by teachers' beliefs is classroom assessment. This study sought to examine the structure of Filipino teachers' beliefs, meanings, propositions, rules, and mental images, more generally referred to as conceptions about assessment. A total of 391 Filipino teachers responded to the Conceptions of Assessment Inventory (COA-III; Brown, 2004). A combination of confirmatory and exploratory factor analyses revealed that the structure of Filipino teachers' conceptions of assessment may be explained by three dimensions: (1) assessment as a means to improve teaching and learning; (2) assessment as a means to hold schools accountable; and (3) assessment as irrelevant to teaching and learning. Implications for pre-service and in-service teacher training and further research are discussed.

Keywords: Conceptions of assessment, beliefs about assessment, Filipino teachers

Introduction

It is generally accepted that teachers' knowledge, beliefs, and thinking about components of the teaching-learning process influence their classroom practices (Kane, Sandretto, & Heath, 2002). What teachers know and believe play a significant role in how they interpret new information and experience, and these interpretations, in turn, guide their instructional practices (Phipps & Borg, 2009). To clarify the overlapping constructs of teacher knowledge, beliefs,

and thinking, Thompson (1992, p. 130) proposed the term “conceptions” to refer to “a more general mental structure, encompassing beliefs, meanings, conceptions, propositions, rules, mental images, preferences, and the like.” Thus, conceptions are proposed to be organizing frameworks through which individuals understand, respond to, and interact with a phenomenon (Brown, 2004).

One important component of the teaching-learning process that is guided by teachers’ conceptions is classroom assessment. Nitko and Brookhart (2007) define assessment as the process of obtaining information using various methods and interpreting these information to make decisions about students. Assessment serves many important purposes in the learning process. When done correctly, it may be used to certify student learning, improve learning and teaching, and even help students to evaluate their own work (Shepard, 2000). Thus, teachers’ conceptions about assessment must be examined because they strongly influence how teachers design, implement, and interpret (the results of) classroom assessments.

Teachers’ Conceptions of Assessment

Brown (2004) proposed four general conceptions of assessment based on its perceived purpose: (1) Assessment improves learning and teaching; (2) Assessment makes students accountable for learning; (3) Assessment is used to hold schools and teachers accountable; and (4) Assessment is irrelevant to teaching and learning. The first three conceptions of assessment were identified by Brown (2002) through literature review. The last conception was proposed by Brown (2002) after noticing that many teachers feel that assessment, particularly standardized assessment, can be detrimental to teacher autonomy and professionalism, and can disrupt student learning.

To test this four-dimensional framework of teachers’ conceptions of assessment, Brown (2004) developed a self-report attitude inventory that measures teachers’ conceptions of assessment using a six-point agreement rating scale. The original instrument consisted of 65 items, but was later trimmed to 50 items after exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The resulting Conceptions of Assessment (COA-III) inventory was first tested among New Zealand primary school teachers where it was found that a four-factor model best represented teachers’ conceptions of assessment. The four-factor model has also been confirmed with Australian primary and secondary school teachers (Brown, Lake & Matters, 2011). Attempts to test the conceptions of assessment model with teachers from other countries have

shown that this four-factor structure is not invariant. A study conducted by Barnes, Fives & Dacey (2017) showed that the four-factor model did not adequately represent US K-12 teachers' conceptions of assessment. Instead, exploratory factor analysis showed that US teachers have three dominant conceptions of assessment - assessment as valid for accountability, assessment as a tool for improving teaching and learning, and assessment as irrelevant (Barnes et al., 2017). Similarly, Fletcher, Meyer, Anderson, Johnston, and Rees (2012) tested the conceptions of assessment model with higher education faculty from four New Zealand universities, and found that a two-factor higher-order model was a better fit compared to Brown's (2004) original four-factor model. Their analysis showed that higher education faculty generally hold either a positive conception of assessment that focuses on improvement or a negative conception of assessment that considers assessment to be irrelevant to their practice.

Departing from a quantitative approach to examining conceptions of assessment, Remesal (2011) did a qualitative study with primary and secondary teachers in Barcelona, Spain. The study examined teachers' conceptions about assessment in terms of four aspects: the learning process, the teaching process, accreditation of learning, and accountability of teachers. Remesal (2011) proposed a four-dimensional bipolar model of conceptions of assessment. In this model, teachers' conceptions of assessment fall into two categories: the pedagogical-regulation pole, which focuses on the monitoring of teaching and learning, and the societal-accreditation pole, which focuses on teachers' accountability and certification of achievement. Remesal (2011) noted, however, that these conceptions do not appear to be mutually exclusive, but may occur in different combinations (extreme pedagogical, mixed pedagogical, mixed societal, and extreme societal). This suggested that teachers may hold contradictory beliefs about how assessment affects teaching and learning and that their conceptions of assessment may not be organized as neatly as Brown's four-factor model (2004).

Classroom Assessment in the Philippines

The Department of Education (2015) emphasized that classroom assessment is an important component of curriculum implementation as it allows teachers to track and measure students' progress and to adjust instruction appropriately. The department's policies and guidelines on classroom assessment are articulated in the DepEd Order No. 8, which has been implemented since School Year 2015-2016. In this document, classroom

assessment is classified as either formative or summative. Formative assessment is defined as either “assessment FOR learning so teachers can make adjustments in their instruction” or as “assessment AS learning wherein students reflect on their own progress” (Department of Education, 2015, p. 3). The memo further described formative assessment as “characteristically informal” and “intended to help students identify strengths and weaknesses in order to learn from the assessment experience” (Department of Education, 2015, p. 3). Their description also emphasized that formative assessment is something that the teacher gives to the student at any time during the teaching and learning process in order to track student progress and make informed instructional decisions. Summative assessment, on the other hand, is described as “assessment OF learning, which occurs at the end of a particular unit” (p. 3) to judge whether or not students have met the content and performance standards specified in the K-12 curriculum (Department of Education, 2015). The memo also noted that the results of summative assessments are recorded and reported to the learners, their parents or guardians, and to school administrators.

It seems that the Department of Education views assessment as serving dichotomous functions - formative, which focuses on using assessment to improve learning and teaching, and summative, which certifies student learning. Remesal (2011) noted that a dichotomous perspective of the functions of assessment is limited and does not sufficiently capture the complexity of the assessment process. Moreover, the Department of Education’s description of formative assessment as a process that teachers perform upon their students fails to recognize the role and responsibility of the students in the learning process. According to Black and Wiliam (2009), formative assessment must include the learners as owners of their learning and as instructional resources for each other. Thus, Filipino teachers might hold conceptions of assessment that are narrow or limited, considering that the Department of Education has explicitly stated such views in their policy guidelines. This notion, however, needs to be verified as there seems to be a dearth of studies on assessment beliefs and practices of Filipino teachers.

One such study examined the assessment literacy of Filipino pre-service and in-service teachers (Balagtas, Dacanay, Dizon, & Duque, 2010). Using a self-report survey, the study found that Filipino pre-service and in-service teachers are weak in terms of the following competencies: administering, scoring, and interpreting results of externally-produced and teacher-made tests; using assessment results to make different types of academic decisions; developing valid student grading procedures; communicating assessment results to students, parents, and other stakeholders; and recognizing unethical, illegal,

and other inappropriate methods and uses of assessment. Furthermore, the study found that teachers are unprepared to use alternative assessments, as they were not trained to do so.

The Present Study

Given that teachers' beliefs are believed to influence classroom practices and outcomes, this study attempts to examine Filipino teachers' conceptions of assessment while addressing the limited literature on classroom assessment in the Philippine context. This is an initial study that seeks to test the structure of the Conceptions of Assessment Inventory (COA-III; Brown, 2004) among Filipino teachers. Previous studies on conceptions of assessment have shown that the factor structure of the COA-III varies across samples (e. g., Barnes et al., 2017; Brown, 2004; Brown & Remesal, 2012; Fletcher et al., 2012). The present study seeks to determine if Filipino teachers' conceptions of assessment may be represented adequately by the model of Brown (2004). If the model is found to be appropriate for Filipino teachers, then the model can be used to inform professional development programs and assessment decisions in the classroom.

Method

Participants

A total of 391 Filipino teachers (74% females, 26% males) participated in the present study. Eighty-seven percent ($n = 342$) teach basic education, and 13% ($n = 49$) teach in the tertiary level. The mean age of the respondents is 32.89 years, with a standard deviation of 10.94 years. On the average, the participants' length of teaching experience is 9.07 years ($SD = 9.23$ years). Majority ($n = 330$ or 84%) of the participants have obtained a Bachelor's degree, but only a handful have completed higher degrees ($n = 56$ or 14% have completed their Master's degrees and $n = 5$ or 2% hold a Doctorate degree). The participants teach various subjects: Mathematics ($n=63$ or 16%), Science ($n=59$ or 15%), English ($n=56$ or 14%), different subjects in self-contained classes ($n=55$ or 14%), professional courses ($n=42$ or 11%), Filipino ($n=24$ or 6%), Social Studies ($n=24$ or 6%), Music, Arts, PE and Health (MAPEH) ($n=21$ or 5%), Christian Living/Religious Education ($n=17$ or 4%), and Technology

and Livelihood Education/Computer (n=10 or 3%). There were 20 teachers (5%) who did not indicate the subjects they are teaching.

Instrument

Filipino teachers' beliefs about assessment were measured using the Conceptions of Assessment Inventory III (COA-III; Brown, 2004). The COA-III is a self-report attitude inventory that measures teachers' degree of agreement or disagreement with statements related to assessment. The scale consists of 50 items that load on two first-order factors (School Accountability and Student Accountability), and on two second-order factors (Improvement, with four subscales, and Irrelevance, with three subscales). The COA-III subscales, their key premises, and sample items are shown in Table 1. The COA-III uses a 6-point positively-packed agreement rating scale (strongly disagree, mostly disagree, slightly agree, moderately agree, mostly agree, strongly agree), which has been shown to be appropriate when respondents are likely to hold positive attitudes toward the construct being measured (Brown, 2004).

Data Analysis

Means and standard deviations were calculated for the nine subscales of the COA-III. Measures of skewness and kurtosis were also computed to provide more information about how the subscale scores were distributed. Cronbach's alpha values were also calculated for each of the subscales to establish internal consistency reliability.

The factor structure of the COA-III was tested using a combination of exploratory and confirmatory factor analyses. First, confirmatory factor analysis (CFA) using IBM SPSS AMOS was used to test Brown's (2004) original four-factor model with two first-order factors (school accountability and student accountability) and two second-order factors (improvement and irrelevance). The following fit indices and criteria were used to establish model fit: Chi-square index is statistically non-significant; the root-mean-square-error-of-approximation (RMSEA) is .06 or less; standardized root mean square residual (SRMR) is .08 or less; and the comparative fit index (CFI) and the Tucker-Lewis index (TLI) are at least .95 (Hu & Bentler, 1999).

Table 1
COA-III Subscales and Sample Items

First-Order Factor	Second-Order Factor	Key Premise	Number of items
School Accountability		Assessment can be used to account for a teacher's, a school's, or a system's use of society's resources.	6
Student Accountability		Assessment is used to hold students accountable for their learning.	7
Improvement	Describe	Assessment is used to describe the abilities, knowledge, and thinking of students.	6
Improvement	Student Learning	Assessment improves student learning.	7
Improvement	Validity	Assessment information is valid.	5
Improvement	Teaching	Assessment improves teaching.	6
Irrelevance	Bad	Assessment is bad for teaching.	5
Irrelevance	Ignore	Teachers may use assessment, but they ignore it.	5
Irrelevance	Accurate	Assessment is inaccurate.	3

Considering that previous studies (e. g., Barnes et al., 2017; Brown & Remesal, 2012; Fletcher et al., 2012) yielded findings that were not consistent with Brown's (2004) four-factor second-order model of conceptions of assessment, the present study also followed an exploratory approach. The purpose of the exploratory factor analysis (EFA) was to discover the underlying dimensions of Filipino teachers' conceptions of assessment without specifying a priori dimensions. This approach was used by Brown and Remesal (2012) in examining the structure of Spanish teachers' conceptions of assessment.

Exploratory factor analysis was conducted using SPSS. Principal axis factoring (PAF) with promax rotation was used to extract the factors that would best represent Filipino teachers' conceptions of assessment. PAF is recommended when the purpose of the analysis is not simply to reduce data, but to express the relationships among the items in a scale in terms of their underlying latent dimensions (Floyd & Widaman, 1995). Also, Fabrigar, Wegener, MacCallum, & Strahan (1999) noted that PAF may be used even when the assumption of multivariate normality is violated. Four criteria were used for deciding on how many factors to retain: (1) Kaiser-Guttman criterion (i.e., retain factors with eigenvalues greater than 1); (2) the scree test (i.e., number of factors to retain is determined by locating the point in the scree plot at which the slope is zero); (3) factors that have at least three items with factor loadings greater than 0.3; and (4) interpretability of the resulting factor structure (Fabrigar et al., 1999). The obtained factor structure was then rotated using an oblique rotation method (i. e., promax rotation) because teachers' conceptions of assessment are thought to have dimensions that would be related to each other. The underlying dimensions found using exploratory factor analysis were then compared with the four-factor model proposed by Brown (2004).

Results

Descriptive Statistics

The mean, standard deviation, skewness, and kurtosis of each COA-III subscale were calculated to determine the distribution of scores. Cronbach's alpha values of each subscale were also determined. These statistics are all presented in Table 2.

Table 2

Descriptive Statistics for the Conceptions of Assessment Inventory (COA-III) (n = 391)

COA-III Subscales	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	Cronbach's α
School Accountability	4.28	0.98	-0.53	0.01	0.87
Student Accountability	4.19	0.88	-0.69	0.62	0.77
Improvement	4.58	0.88	-1.26	1.91	0.95
Describe	4.83	1.08	-0.48	5.79	0.86
Student Learning	4.79	0.97	-1.35	1.99	0.89
Validity	4.12	0.90	-0.65	0.30	0.71
Teaching	4.61	0.97	-0.70	1.85	0.84
Irrelevance	3.83	0.82	-0.56	0.66	0.81
Bad	4.13	1.03	-0.31	0.00	0.64
Ignore	4.45	1.20	-0.89	0.34	0.83
Accurate	2.93	0.93	0.25	0.55	0.37

The rating scale used for the COA-III is a six-point positively-packed agreement rating scale, with mean scores ranging from 2.93 to 4.83. The standard deviations show minimum dispersion of scores from the mean of each subscale. The skewness values range from -1.26 to 0.25, but most of the values are negative, indicating that the scores for the subscales are concentrated on the higher values. Subscales whose skewness values are between -0.5 to 0.5 indicate that the distribution of scores is slightly skewed, while those whose skewness values are smaller than -1.0 or larger than 1.0 indicate a distribution that is moderately skewed. This is to be expected since the scale that was used is positively packed (i.e., two negative responses and four positive responses). Positively packed scales have been shown to be effective in generating a variety of responses when participants are inclined to respond positively to items (Brown, 2004). The positive kurtosis values indicate that the distribution is leptokurtic (i.e., more peaked than the normal distribution and has fatter tails).

Confirmatory Factor Analysis

Brown (2004) proposed that the structure of teachers' conceptions of assessment may be represented by a four-factor model with two first-order

factors (School Accountability and Student Accountability) and two second-order factors (Improvement and Irrelevance). This four-factor second-order model was tested among a heterogenous group of Filipino teachers.

All of the items in Brown's (2004) COA-III loaded significantly on their hypothesized factor ($p < .05$), except for two items under Irrelevance-Accurate (i.e., *Assessment is an imprecise process* and *Assessment results should be treated cautiously because of measurement error*) which had non-significant factor loadings. These two items had standardized factor loadings below .4, which indicates that these items seem to be unrelated to the underlying dimension they are supposed to measure. Table 3 shows the standardized factor loadings and standard errors of each item in the COA-III.

Table 3
COA-III Statements, Factor Loadings, and their Standard Errors

Factors and Statements	Unstandardized Factor Loading	Standardized Factor Loading	Standard Error
<u>School Accountability</u>			
Assessment measures the worth or quality of schools	1.000	.654	
Assessment is an accurate indicator of a school's quality	1.155	.764	.089
Assessment shows the value schools add to student learning	.924	.775	.070
Assessment is a good way to evaluate a school	1.126	.803	.083
Assessment influences the way teachers think	1.085	.806	.080
Assessment keeps schools honest and up-to-scratch	.664	.514	.072
<u>Student Accountability</u>			
Assessment selects students for future education or employment opportunities	1.000	.487	
Assessment is comparing student work against set criteria	.926	.467	.128
Assessment determines if students meet qualifications standards	1.410	.838	.142

Factors and Statements	Unstandardized Factor Loading	Standardized Factor Loading	Standard Error
Assessment is assigning a grade or level to student work	1.108	.606	.130
Assessment places students into categories	1.172	.621	.136
Assessment is checking off progress against achievement objectives	1.096	.620	.127
Assessment is completing checklists	.538	.258	.118
<u>Improvement-Describe</u>			
Assessment is a way to determine how much students have learned from teaching	1.000	.790	
Answers to assessment show what goes on in the minds of students	.902	.716	.058
Assessment measures students' higher order thinking skills	1.048	.809	.057
Assessment establishes what students have learned	1.015	.845	.052
Assessment identifies student strengths and weaknesses	1.100	.879	.054
Assessment identifies how students think	.810	.273	.150
<u>Improvement-Student Learning</u>			
Assessment feeds back to students their learning needs	1.000	.769	
Assessment helps students improve their learning	.967	.809	.055
Assessment is appropriate and beneficial for children	1.001	.815	.057
Assessment provides feedback to students about their performance	1.001	.862	.053
Assessment is an engaging and enjoyable experience for children	.636	.508	.062
Assessment makes students do their best	.948	.767	.058
Assessment is a positive force for improving social climate in a class	.758	.589	.063
<u>Improvement-Validity</u>	1.000	.695	

Factors and Statements	Unstandardized Factor Loading	Standardized Factor Loading	Standard Error
Assessment results are trustworthy			
Assessment results predict future student performance	1.020	.621	.095
Assessment results are consistent	.539	.327	.093
Assessment is objective	.887	.575	.089
Assessment results can be depended on	.928	.610	.088
<u>Improvement-Teaching</u>			
Assessment influences the way teachers think	1.000	.608	
Assessment is integrated with teaching practice	1.287	.842	.099
Assessment changes the way teachers teach	1.172	.682	.104
Assessment information modifies ongoing teaching of students	1.134	.712	.098
Assessment allows different students to get different instruction	1.238	.507	.140
Assessment information is collected and used during teaching	1.257	.695	.110
<u>Irrelevance-Bad</u>			
Teachers pay attention to assessment only when stakes are high	1.000	.496	
Assessment interferes with teaching	.646	.278	.139
Teachers are over-assessing	1.197	.671	.140
Assessment is unfair to students	1.305	.721	.147
Assessment forces teachers to teach in a way against their beliefs	.709	.344	.128
<u>Irrelevance-Ignore</u>			
Teachers ignore assessment information even if they collect it	1.000	.729	
Assessment has little impact on teaching	.893	.616	.078
Teachers conduct assessments but make little use of the results	.842	.635	.071
Assessment is value-less	.977	.752	.070

Factors and Statements	Unstandardized Factor Loading	Standardized Factor Loading	Standard Error
Assessment results are filed and ignored	1.076	.797	.073
<u>Irrelevance-Accurate</u>			
Teachers should take into account the error and imprecision in all assessment	1.000	1.618	
Assessment is an imprecise process	.021	.028	.079
Assessment results should be treated cautiously because of measurement error	.152	.246	.518

It should also be noted that four first-order factors (Describe, Student Learning, Validity, and Teaching) had significant factor loadings (standardized β 's $> .85$) on their second-order factor (i.e., Improvement). However, of the three first-order factors (Bad, Ignore, and Accurate), only two factors (Bad and Ignore) loaded significantly (standardized β 's $> .90$) on their second-order factor (i.e., Irrelevance), indicating that the first-order factor Irrelevance-Accurate may be unrelated to its hypothesized second-order factor.

Correlations among the second-order factors are shown in Table 4. Large (r 's $> .50$) effect sizes were observed among school accountability, student accountability, and improvement. Large effect sizes (r 's $> .50$) indicate that the proportion of variance shared by school accountability, student accountability, and improvement is at least 25%. Take for instance, school accountability and student accountability have an observed $r = .73$, which indicates that 53.29% of the variation in school accountability may be explained by the variation in student accountability. This simply means that differences in Filipino teachers' beliefs pertaining to assessment as a means for ensuring school accountability may be accounted for by the differences in their beliefs regarding assessment as a means to promote student accountability. This is not to say that the two factors are similar. Rather, these factors vary together, and in the same direction, suggesting that when teachers' hold positive beliefs about assessment as necessary for school accountability. They also tend to hold positive beliefs about assessment as important for ensuring student accountability. The observed positive relationships among these factors also provide evidence for convergent validity.

Moreover, the observed indirect relationship between school accountability and irrelevance ($r = -.18$), and between student accountability and irrelevance ($r = -.23$) provide support for discriminant validity. This means that if teachers think of assessment as helpful for holding schools and students accountable, then they most likely will not consider assessment as irrelevant. However, the observed effect sizes ($r^2 < .10$) were small, which means that the proportion of variance in school accountability and student accountability that may be explained by the variation in irrelevance is less than 5%. It should also be noted that the observed effect size ($r^2 = .01$) between improvement and irrelevance is not significant, which means that these factors are unrelated.

Table 4
COA-III Second-Order Factor Correlations

Factor	(1)	(2)	(3)	(4)
(1) School Accountability	-			
(2) Student Accountability	.73*	-		
(3) Improvement	.76*	.75*	-	
(4) Irrelevance	-.18*	-.23*	.01	-

* Correlation is significant at the 0.05 level (2-tailed)

The four-factor second-order model of the COA-III showed poor fit: $\chi^2 = 3,617.089$, $df = 1162$, $p\text{-value} = .000$; $\chi^2 / df = 3.113$; RMSEA = 0.074; SRMR = 0.1179; CFI = 0.784; TLI = 0.772. A generally accepted procedure in CFA is to respecify the model by removing items that have low factor loadings or by allowing error terms to be correlated. This modified model is then tested (see Byrne, 2010 for an illustration). This procedure is more appropriate, however, if the second (i.e., respecified) model were to be tested on a sample that is different from the one that was used to derive the first model (Sharma, 1996). A common practice is to use a relatively large sample, split it into two subsamples – a derivation sample and a cross-validation sample, and use the cross-validation sample to test the modified model (Floyd & Widaman, 1996). In this study, however, the obtained sample size ($n = 391$), was adequate only for deriving a model, but not for a cross-validation. Thus, we opted not to respecify and test a second model at this time for lack of an adequate cross-validation sample.

The obtained fit indices indicate that Filipino teachers' responses on the COA-III did not sufficiently support the four-factor model of conceptions of

assessment proposed by Brown (2004). This also suggests that the Filipino teachers may hold different conceptions of assessment that are not captured by Brown's model. The results of the CFA also lend support to our initial proposal of doing an EFA in order to discover the underlying structure of Filipino teachers' conceptions of assessment.

Exploratory Factor Analysis

Previous research has shown that the factor structure of the COA-III (Brown, 2004) varies across samples from different countries (Barnes et al., 2017; Brown & Remesal, 2012; Fletcher et al., 2012). There is empirical evidence, therefore, to support the notion that the structure of Filipino teachers' conceptions of assessment may be different from that of New Zealand teachers who composed the sample for the original inventory. Thus, the current study explored the underlying dimensions of Filipino teachers' conceptions of assessment without specifying a priori the structure of these dimensions.

To determine whether the data are adequate for an EFA, we first examined the correlations among the 50 items of the COA-III. We found moderate correlations, indicating that the items may be grouped into homogeneous factors that measure the similar underlying dimensions. Second, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was found to be .948. According to Sharma (1996), a KMO value of .90 or higher means that the items are homogeneous enough and, therefore, appropriate for factoring. Third, Bartlett's test of sphericity resulted to a p-value less than $\alpha = .05$, which indicates that there is redundancy between items that can be summarized into factors, and therefore factor analysis is appropriate (Sharma, 1996).

Exploratory factor analysis using principal axis factoring with oblique (promax) rotation revealed four factors with eigenvalues greater than one. The scree plot, however, showed that there are three factors above the inflection point, which suggests that only three factors ought to be retained based on this heuristic. Since the two criteria (i.e., Kaiser-Guttman and scree plot) resulted to inconsistent results, we examined the items that loaded significantly on the extracted factors. Closer inspection of the factor loadings showed that for the fourth factor, there were only two items that had factor loadings greater than .3, which suggests that the fourth factor may no longer be interpretable. Thus, we decided that only three factors would be retained. These three factors explained 45.45% of the variance. Table 5 shows the eigenvalues, the percentage of values accounted for by each factor before and after rotation, and the number of items in each factor that had factor loadings of 0.3 or higher.

Table 5
Eigenvalues, Variance Explained and Items Included for the Three-Factor Model

Factor	Eigenvalues	Percentage of Variance Explained		No. of Items
		Unrotated	Rotated	
		1	16.878	
2	5.450	10.900	9.963	8
3	1.678	3.356	2.464	6

It should be noted that the items did not load on the original factors conceptualized by Brown (2004). Also, only 31 items were retained from the original scale because these were the items that had factor loadings greater than 0.3. We then examined the resulting factor structure to see if it was interpretable and theoretically sensible, as suggested by Fabrigar et al. (1999). Table 6 shows the COA-III items that were retained for a three-factor solution, their factor loadings, and the Cronbach's alpha coefficients of the new subscales.

There were 17 items that loaded on the first factor, which explained 33.023% of the variance after rotation. Fifteen of these items measured the Improvement function of assessment in the original COA-III inventory. Two items focused on Student Accountability (*Assessment determines if students meet qualifications standards* and *Assessment is checking off progress against achievement objectives*). These items, however, seem to be still related to the notion that assessment serves an improvement function, because it also implies that they have improved when students meet qualification standards or show progress against achievement objectives. Thus, we named Factor 1 as the *Improvement* factor, similar to how Brown's (2004) and Barnes et al.'s (2017) interpretation. Since there were 17 items that loaded on this factor, we also checked if there would be a second-order factor by running another EFA using the 17 items. Only one factor was extracted and it explained 59.043% of the variance, suggesting unidimensionality. Thus, we retained all 17 items under the Improvement factor.

Table 6

Factor Loadings of the Oblique, Three-Factor, 31-item Solution with Internal Consistency Coefficients

Factor / Item	Factor			Cronbach's α
	1	2	3	
<u>Factor 1 – Improvement</u>				0.951
21. Assessment provides feedback to students about their performance	0.966			
22. Assessment identifies student strengths and weaknesses	0.935			
45. Assessment helps students improve their learning	0.812			
12. Assessment feeds back to students their learning needs	0.798			
26. Assessment establishes what students have learned	0.774			
48. Assessment is a way to determine how much students have learned from teaching	0.77			
35. Assessment is integrated with teaching practice	0.758			
20. Assessment determines if students meet qualifications standards	0.733			
41. Assessment influences the way teachers think	0.712			
6. Assessment information is collected and used during teaching	0.664			
32. Assessment measures students' higher order thinking skills	0.635			
38. Assessment is appropriate and beneficial for children	0.593			
9. Assessment makes students do their best	0.582			
28. Assessment is checking off progress against achievement objectives	0.566			
18. Assessment information modifies ongoing teaching of students	0.534			
33. Assessment changes the way teachers teach	0.522			
7. Assessment allows different students to get different instruction	0.41			
<u>Factor 2 – Irrelevance</u>				.862
43. Teachers conduct assessments but make little use of the results		0.756		
23. Teachers ignore assessment information even if they collect it		0.721		
50. Assessment results are filed and ignored		0.696		

8. Teachers are over-assessing	0.682
1. Teachers pay attention to assessment only when stakes are high	0.641
46. Assessment is value-less	0.595
40. Assessment has little impact on teaching	0.525
11. Assessment is unfair to students	0.502

Factor 3 – School Accountability

0.853

42. Assessment provides information on how well schools are doing	0.835
37. Assessment is a good way to evaluate a school	0.810
30. Assessment is an accurate indicator of a school's quality	0.760
17. Assessment measures the worth or quality of schools	0.448
27. Assessment places students into categories	0.378
24. Assessment results predict future student performance	0.333

The second factor had eight items and explained 9.963% of the variance after rotation. Items in this factor pertain to the notion that assessment is irrelevant to the teaching and learning process; thus, we named the factor *Irrelevance*, similar to how Brown (2004) and Barnes et al. (2017) interpreted the same results. Finally, the third factor contained five items that describe assessment as necessary for holding schools accountable and one item that describe assessment as a means for predicting future student performance, which is related to the notion that assessment information is valid. However, this item may still be considered as related to the school accountability function of assessment. It also indicates that the school has been doing well in making good use of its resources when students perform well into the future. Thus, we labeled the third factor as *School Accountability*. This factor accounted for 2.464% of the variance after rotation.

To verify this three-factor model, we ran a second EFA with a fixed three-factor solution. The second EFA yielded a three-factor model that explained 44.90% of the variance and had a factor structure similar to the initial model. The first factor that was extracted in the second EFA included items that were related to the *Improvement* function of assessment. Sixteen of the 17 items that loaded on the first factor in the initial EFA also loaded highly on the first factor in the second EFA (i.e., fixed three-factor solution). The second factor extracted in the second EFA had items that were related to the *School Accountability* function of assessment. In the first EFA, the School Accountability

factor was the third factor extracted and had six items. Of these six items, five also loaded highly on the Accountability factor of the second EFA. The third factor extracted in the second EFA included items related to the *Irrelevance* conception of assessment. These same items also loaded on the *Irrelevance* factor in the first EFA. The results of the second EFA with a fixed three-factor solution provide tentative support for our decision to retain three factors in the initial solution. We recognize, however, that since we used the same data set to run a second EFA, our findings need to be verified using an independent sample. Nonetheless, the second EFA suggests that a three-factor model of conceptions of assessment may be more functional for Filipino teachers than Brown's (2004) original four-factor model.

To provide evidence of the functionality of the three-factor solution, we also examined factor intercorrelations. As expected, the Improvement and School Accountability factors are significantly correlated and the effect size is large ($r > .5$). Also, there is no significant relationship between the Irrelevance and School Accountability factors. These results are consistent with the findings of Brown (2004).

Table 7
Correlations among the Three Factors of the COA-III

Factor	(1)	(2)	(3)
(1) Improvement	-		
(2) Irrelevance	.322**	-	
(3) School Accountability	.679**	-0.028	-

**Correlations are significant at the .001 level. (2-tailed)

An interesting result is the positive and significant relationship between Improvement and Irrelevance ($r = .322$). This seems to be counter intuitive because when assessment is seen to improve learning and teaching, then it should be considered as relevant, as explained by Brown (2004). However, it is also likely that, although teachers conceive of assessment as something that is helpful in improving learning and teaching, they may still think of some aspects of assessment as irrelevant, particularly when assessment results are not put to good use. An illustration of this mixed conceptions of assessment may be seen in a qualitative study of Spanish teachers' conceptions of assessment conducted by Remesal (2011).

Remesal (2011) found that teachers's views about assessment belong to two types: pedagogical (focusing on the monitoring of teaching and learning) and societal (focusing on teachers' accountability and the certification of achievement). However, Remesal (2011) pointed out that these two sets of beliefs are not mutually exclusive, and that teachers may hold a combination of these beliefs – a purely pedagogical conception, a purely societal conception, a mixed pedagogical conception, and a mixed societal conception. Thus, the observed positive and significant relationship between the Irrelevance and Improvement factors provide support for Remesal's (2011) contention that a strictly dichotomous interpretation of the functions of assessment (positive vs. negative or formative vs. summative) may not be functional or realistic. Teachers may indeed hold contradictory beliefs about the importance and relevance of assessment in teaching and learning. This means that while teachers believe that assessment is important in the teaching and learning process, current practices in and expectations from the educational system may lead them to also think of assessments as irrelevant. For instance, teachers in basic education may use assessments to facilitate student learning and enhance classroom instruction. In this case, assessment serves an improvement function. However, these same teachers also need to prepare their students to take high-stakes tests, such as the National Achievement Test prescribed by the Department of Education. Thus, classroom assessments may no longer serve a pedagogical function when they are meant only to prepare students to get high scores in standardized tests, which are often used to rank schools, particularly among public schools. Private schools, on the other hand, may require teachers to develop a variety of assessments in compliance with the Department of Education (2015) policy guidelines on classroom assessment. The policy guidelines prescribe specific types of assessments (i.e., written works, performance tasks, and quarterly assessment) that teachers must include in rating student performance, whether or not these assessments are aligned with the learning outcomes or with teaching and learning activities. When teachers develop assessment tasks out of compliance, they might consider assessment to be irrelevant or unimportant to the teaching and learning process.

The situation in higher education may not be any different from that of the basic education sector. Since higher education is intended to prepare students for a particular profession, there is often a great deal of pressure for university professors to cover the necessary content that students need to know. Thus, assessments in the university level may be important only as far as ensuring that students have gained the necessary knowledge and skills so that they can be ready to take on roles related to their chosen profession. The

pedagogical purpose of assessment (i.e., to help students improve their learning, to facilitate student engagement; to improve classroom instruction) may be overlooked when the focus of assessment is to certify student achievement of desired knowledge and skills.

Discussion

The current study tested the factor structure of the COA-III inventory (Brown, 2004), an instrument designed to measure teachers' conceptions about the functions of assessment. Using a combination of CFA and EFA, we found support for findings of previous studies about the lack of invariance of the scale across samples from different countries (see Brown & Remesal, 2012; Fletcher et al., 2012; and Barnes et al., 2017). We also found that Filipino teachers may hold contradictory beliefs about the functions of assessment.

These findings have implications for the training of both preservice and in-service teachers in assessment. Teachers' beliefs strongly influence their actual classroom practice, therefore, it is important for teacher education institutions to incorporate in their assessment courses some opportunities to examine their beliefs. Teacher educators need to address students' misconceptions or faulty understandings of assessment. In-service training of teachers must also include sessions on assessment that would focus not just on the strategies for classroom assessment, but also on examining teachers' beliefs about the purpose and functions of assessment.

Considering that the structure of Filipino teachers' conceptions of assessment seems to be misrepresented by the current model of Brown (2004), it would benefit teacher educators, school administrators, policy makers, and the teachers themselves if a more indigenous model would be developed. Remesal (2011) pointed out that the assessment beliefs of teachers may be affected by the structure of the educational system. This means that in developing an indigenous model of Filipino teachers' conceptions of assessment, their context, experiences, and challenges within the educational system must also be considered. Future research, therefore, may be undertaken using either a qualitative or a mixed-methods approach to determine a model that would more closely represent their conceptions of assessment. Eventually, such model may be used to inform teacher training programs and to examine assessment practices in the classroom.

Finally, our study is an attempt to address the dearth of research on classroom practices of Filipino teachers. It is our hope that this and future

research will not only help Filipino teachers in improving their practice, but will also give them voice in the literature on educational assessment.

References

- Balagtas, M. U., Dacanay, A. G., Dizon, M. A., & Duque, R. E. (2010). Literacy level on educational assessment of students in a premiere teacher education institution: Basis for a capability building program. *The Assessment Handbook, 4*, 1-19.
- Barnes, N., Fives, H., & Dacey, C. M. (2017). US teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education, 65*, 107-116.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*, 5-31.
- Brown, G. T. L. (2002). *Teachers' conceptions of assessment* (Auckland, NZ, University of Auckland).
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*, 301-318.
- Brown, G. T. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological reports, 99*(1), 166-170.
- Brown, G. T., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology, 15*, 75-89.
- Brown, G. T., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*, 210-220.
- Department of Education. (2015). *Policy guidelines on classroom assessment for the K-12 Basic Education Program*. Manila, Philippines.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.
- Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and students conceptions of assessment in higher education. *Higher Education, 64*, 119-133.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6, 1-55.
- Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, 72(2), 177-228.
- Nitko, A. J., & Brookhart, S. M. (2007). Educational assessment of student. Englewood Cliffs, NJ: Merrill Prentice Hall.
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37, 380-390.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27, 472-482.
- Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley & Sons.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 4-14.
- 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: A project of the National Council of Teachers of Mathematics* (pp. 127-146). New York, NY, England: Macmillan Publishing Co, Inc.