



Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Version

This is the Accepted Manuscript version of the following article. This version is defined in the NISO recommended practice RP-8-2008

<http://www.niso.org/publications/rp/>

Suggested Reference

Brown, G. T., & Remesal, A. (2017). Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian teachers. *Studies in Educational Evaluation*, 55, 68-74. doi: [10.1016/j.stueduc.2017.07.003](https://doi.org/10.1016/j.stueduc.2017.07.003)

Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

This is an open-access article distributed under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivatives](https://creativecommons.org/licenses/by-nc-nd/4.0/) License.

For more information, see [General copyright](#), [Publisher copyright](#), [SHERPA/RoMEO](#).

Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian secondary
teachers

Gavin T. L. Brown

The University of Auckland

Ana Remesal

Universitat de Barcelona

Correspondence concerning this manuscript should be addressed to Prof. Gavin T. L. Brown,
School of Learning, Development, and Professional Practice, Faculty of Education and Social
Work, The University of Auckland, Private Bag 92019, Auckland, 1142, New Zealand or by
email to: gt.brown@auckland.ac.nz

Highlights

- Survey of 563 Ecuadorian primary and secondary school teachers with 2 inventories
- Exploratory, confirmatory factor analyses and structural equation modeling used
- Spanish QMCoA and New Zealand TCoA inventories fit well to acceptably
- Means strongest for improvement, caution, and societal control factors
- Inter-correlations between inventories consistent with high-stakes examination system

This is the author pre-print accepted version of

Brown, G. T. L., & Remesal, A. (2017, accepted). Teachers' conceptions of assessment:
Comparing two inventories with Ecuadorian teachers. *Studies in Educational
Evaluation*.

Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian teachers

Abstract

Teacher beliefs about the purpose of assessment matter to how assessment is implemented in classroom settings. Two different teacher conceptions of assessment inventories have originated separately in Spain and New Zealand. This study examined the extent (a) to which the models for each inventory could be recovered and (b) the relationship between inventories. Responses were obtained from 566 Ecuadorian primary and secondary teachers in two rounds of surveying. Results showed that with a few modifications both the Spanish and New Zealand models fit the data. Mean scores were strongest for improvement, caution, and societal control factors. Factor inter-correlations between the Spain and New Zealand models indicated that the teaching, certifying, and accounting domains were moderately correlated with accountability and improvement purposes, while societal control correlated with caution, and formative regulation correlated with irrelevance. These results are consistent with how teachers seem to conceive of assessment within strong examination systems.

Keywords: assessment; beliefs and conceptions; inventories; teachers; structural equation modelling; Ecuador

The influence of conceptions on the daily school practice is out of question today. In the last decade, different models of how teachers understand the purpose, role, and effects of assessment have been published (Barnes, Fives, & Dacey, 2015). Research has identified that the two predominant purposes of assessment (i.e., formative improvement vs. summative evaluation) create substantial tensions for teachers (Bonner, 2016). A further widely attested phenomenon in teacher thinking about assessment is that these beliefs tend to reflect the social, historical, and cultural priorities established in each jurisdiction in which teachers are employed (Fulmer, Lee, & Tan, 2015). This means that interpreting teacher conceptions of assessment needs to pay attention to contextual factors operating in a jurisdiction with its particular systemic requirements (Remesal, 2007).

While there is conceptual coherence across the various strands of research (e.g., agreement that accountability and improvement are dominant beliefs), there is little empirical evidence comparing and contrasting research instruments from the different research strands. This means that claims of coherence have as yet not been explicitly tested. This paper addresses this need by examining the validity of two published self-report inventories (Brown 2004; Remesal & Brown, 2015) into teacher conceptions of assessment in one particular cultural context (i.e., Ecuador) and attempting to generate a model that captures the interaction between the supposedly similar factors. These two inventories were selected, in part, because they appear to have had substantial influence in how teacher conceptions of assessment have been framed in recent reviews (Barnes, Fives, & Dacey, 2015; Bonner, 2016; Fulmer, Lee, & Tan, 2015). Hence, evaluating the relationship of the two inventories may be of substantial benefit to the field.

Teacher Conceptions of Assessment

Since recent reviews of the teacher conceptions of assessment literature exist (Barnes, Fives & Dacey, 2015; Bonner, 2016; Fulmer, Lee, & Tan, 2015), this section will briefly

outline the most important issues. The *educational* purpose of assessment is to provide information about student progress and needs so as to guide teachers in how they should plan and implement subsequent teaching, while at the same time providing insights to students as to what they should do to improve their own learning (Brown, 2008; Remesal, 2011). In contrast, the *accountability* orientation (Remesal, 2011), sometimes known as *evaluation-orientation* (Brown, 2008), uses assessment to either certify student attainment of expected standards or qualifications (i.e., public examinations for entry to further opportunities), or more generally to identify, so as to reward or punish, highly effective or ineffective, respectively, teachers and/or schools (Nichols & Harris, 2016). The argument simply is that teachers tend to endorse the purposes and functions deemed appropriate by both social norm and official policy. These two purposes or orientations, in accordance with Scriven's (1967) analysis, have been largely captured with the terms *formative* and *summative* assessment, respectively.

Remesal (2007; 2011) presented a qualitative study on primary and compulsory secondary school teachers in Spain (QMCoA, Qualitative Model of Conceptions of Assessment). She reported similar results to Brown, Lake and Matters (2011) in that teachers' conceptions of assessment appeared to be linked to the school level in which teachers worked. More specifically, primary teachers were more concerned with formative assessment while compulsory secondary teachers viewed assessment more as a tool for accountability. It is noteworthy, however, that the school level by itself does not convey the predominant purpose of assessment in each case. Rather the final nature of the educational level within the broader social system and its educational policies defines the purpose. In other words, these results could differ if primary teachers had a social responsibility of declaring pupils' final achievement with a definitive value, as is seen in Hong Kong or China (Brown, Hui, Yu, & Kennedy, 2011).

Remesal's model structures conceptions around how teachers conceive of assessment influencing four different aspects of the teaching and learning process: (1) the act of teaching, (2) the act of learning, (3) the act of providing an accreditation to or certification of learning results, and finally (4) offering accountability to different classroom-external agents, including school administrators, families and policy makers. In her model, any teacher dynamically constructs *conceptions*, that is, *organised beliefs systems*, about each of these four 'spaces of influence of assessment'. Simultaneously this model characterizes the teachers' conceptions as either inclined to a formative-regulatory view or as a non-regulatory accreditative tendency, often with a mixed nature due to the complexity of educational praxis. In other words, in Remesal's model the dilemma lies not on whether the teachers conceive of assessment as improving learning versus giving account, but rather on how they think this improvement should be carried out and monitored and what form the accountability or accreditation should take. This complex mixed approach is necessary since by its own nature the educational system requires inevitably teachers to engage in both basic functions of assessment. This type of model is bifactor in that two different independent dimensions are used to predict responses to items. This bifactor model was validated in a study of foreign language teachers (Spanish as Foreign Language in particular) in school as well as non-formal schooling contexts (Remesal & Brown, 2015). Those Spanish as a Foreign Language teachers agreed most with the formative regulation version of all four conceptions and more-or-less rejected the use of student assessment to evaluate the quality of teaching.

Rather than relying on the bifactor approach, Brown's Teachers Conceptions of Assessment (TCoA) has used a classic simple structure factor model approach to tackle complexity by introducing a hierarchical structure. This model has worked well with primary and secondary teachers in New Zealand (Brown, 2011) and Queensland (Brown, Lake, & Matters, 2011). However, cross-language and societal studies with the TCoA have

consistently shown that the New Zealand developed hierarchical structure of factors fails to replicate. Nonetheless, many items tend to group in identical factors giving partial validity to the items and their factors (Brown & Michaelides, 2011; Brown & Remesal, 2012; Gebril & Brown, 2014).

Despite this, endorsement of items and factors related to assessment for improved teaching and student learning seems to be consistently strongest (Bonner, 2016; Barnes, Fives, & Dacey, 2015). Nonetheless, interesting differences in strength of agreement with four purposes of assessment and the inter-correlations of those purposes seem to align with contextual factors. For example, in the examination-heavy culture of China and Hong Kong (Brown, Hui, Yu, & Kennedy, 2011; Brown, Kennedy, Fok, Chan, & Yu, 2009), assessment for improvement was highly correlated with assessment for student certification or accountability, unlike low-stakes contexts such as New Zealand (Brown, 2011) and Queensland (Brown, Lake, & Matters, 2011). Within jurisdictions it has been consistently reported that secondary school teachers agreed more than primary teachers that assessment was for student accountability while agreeing less that it was for improvement (Brown, 2011; Brown, Lake, & Matters, 2011). These differences suggest that the introduction of national qualifications assessment systems in secondary schooling tend to be associated with different assessment practices and conceptions.

In light of these findings, we might expect that Remesal's QMCoA, having been originally developed in Spanish, would have better fit to Ecuadorian teachers than Brown's TCoA. Nonetheless, we might expect the TCoA items to group in similar ways in Ecuador without replicating the hierarchical structure. We might also expect secondary teachers to have stronger endorsement of student accountability ideas over formative assessment notions.

The school system in Ecuador

Ecuador is a multilingual and multicultural country with slightly over 16,000,000 inhabitants. Over 60% of the population lives in urban areas. There are three main ethnic groups in Ecuador: 'mestizos' (i.e., a person of mixed race, especially of Spanish and American Indian parentage), Afro-Ecuadorians, and Indians. The population of Indians consists of thirteen different ethnic groups, all recognized as individual nationalities in the current constitution (i.e., Awa, Achuar, Chachi, Kichwa, Shuar, Tsáchila, Huaorani (Huaó), Siona, Secoya, Shiwiar, Cofán, Epera, and Záparos).

A series of policy reforms have been successful in reducing functional illiteracy from 46.5% in 1974 to 21.3% in 2001 (Viteri, 2006). The school population has increased from 68.6% in 1982 up to 90.1% in primary school and from 29.5% up to 44.6% in compulsory secondary school in 2001. Given the cultural plurality of Ecuador, great effort in these processes of reforms has focused on the interculturality of the Ecuadorian society (Walsh, 1998). Currently, the Ecuadorian school system follows a 6+4(+2) structure prior to university entrance (i.e., compulsory primary and secondary, followed by optional college).

Torres (2006) shares some recent historical data of the educational state of the art in Ecuador. According to her study, in 2000 Ecuadorians remained 7.5 years in school in average and only 29% of the population finished secondary school. One out of 10 children had to retake first grade, one out of three did not finish primary education, and nine of ten children in rural areas did not finish secondary school. Results in Mathematics and Language in national exams (*Aprendo*) were very low in 1996, and they even decreased by the year 2000.

The government launched in 2006 an immense renewal project of the educational system to fight the low level of school completion and performance, including basic preparation for teaching, in line with other movements in Latin America (Dussel, 2001). New schools with the latest technological resources are being built. Small unitary schools

dispersed in rural areas lacking the minimal infrastructure are being replaced by huge mega-schools that gather pupils from a big area around. These so-called millennium educational units (*Unidad Educativa del Milenio*) receive several thousands of students, often in two or even three shifts a day. The bad reputation of the public school (Ponce, Bedi, & Vos, 2002) is slowly changing nowadays.

Ecuadorian education is governed by the 2011 *Ley Orgánica de Educación Intercultural*, which provides general education up to age 15, when students select either a General or Vocational senior high school orientation (OECD, 2016). The Ecuadorian Ministry of Education (2016) reports satisfying results in 2016: the number of children attending pre-school has increased 10.6 times from 2007 to 2016; in 2016, 96.23% of children between 5 and 12 years of age were attending primary school, while 72.25% of teenagers aged between 12 and 18 years were attending compulsory secondary school, and the national illiteracy ratio decreased 5.65% in 2016.

School in Ecuador is generally characterized by strong traditional conventions. There are two peculiarities in the Ecuadorian school system which we want to underline for the sake of this study: first, the authority role of the teacher within the classroom is still a high cultural value (though this seems to be changing, in line with other social changes across countries and cultures); second, the traditional threshold for passing a course is set at 70% of performance level, as opposed to Spain, for instance, with passing threshold at 50%.

Method

Research Questions

In this current study we administered both inventories to two samples of teachers. Our main research question was: *What beliefs do these Ecuadorian teachers have about assessment?* The second question was: *What is the relationship between the TCoA and the QMCoA?* To address this, first the validity of the original models was evaluated and then the

inter-relationship between inventories was examined. Thus, the conceptions of the teachers are examined in three ways: (1) Remesal's QMCoA, (2) Brown's TCoA, and (3) the joint relationship of the TCoA and QMCoA.

Design

In closing the first semester of a Master's program, carried out as a professional development (PD) project within a national government initiative of educational renewal, 360 secondary mathematics and history teachers taking the course were invited to respond to the questionnaires, QMCoA and TCoA, and provide basic demographic information. The PD program lasted one year divided into two independent semesters with theoretical and praxis-related courses. After the first semester, consisting of an intensive 1-week face-to-face program, followed by an online complementary module, the teachers were invited to respond to the questionnaires online, on a voluntary basis.

Three years later, the same professional development initiative was repeated. This round included 450 primary and secondary school teachers, who were invited to take part in the second round of the survey. While this design relies on a convenience self-selected sample, it does reflect a group of Ecuadorian teachers who are seeking to improve the quality of their teaching.

Participants

Data were obtained from two rounds of surveying. Round 1 (2014) obtained responses from 190 of 358 secondary school mathematics or history teachers participating in the course. Round 2 (2017) obtained responses from 376 of the 470 teachers in the course. Thus, the total response rate was 68%. In Round 2, participating teachers worked at either primary ($n=116$) or secondary ($n=260$) levels. Participants were generally middle-aged (two-thirds were between 35-50 years old), experienced (almost two-thirds had >10 years teaching), and 50%

more were women than men (Table 1). The teachers came in both occasions from all over the country, and gathered together in the capital to take the first face-to-face week of the course.

Table 1. Participant Demographics

Demographic	<i>N</i>	%
Age		
< 35	106	18.7
35 - 50	394	69.6
> 50	66	11.7
Gender		
Men	220	38.8
Women	346	61.1
Teaching Experience		
Less than 2 years	2	0.3
Between 2 - 5 years	68	12.0
Between 5 - 10 years	140	24.7
More than 10 years	356	62.9
Level of Teaching		
Primary	116	20.5
Secondary	448	79.5
Total	566	

In addition to these basic demographic data, teachers were asked to appraise their preparation for assessment duties. One option indicated lack of competence, while three options indicated self-reported competence and an attribution as to whether that arose from initial teacher education, in-service professional development, or through their own self-learning (Table 2). Difference in percent selecting each option between primary and secondary school teachers was not significant ($\chi^2=2.90, p=.41$). Very few reported lacking competence, a quarter attributed their competence to initial teacher education, and just over half attributed their competence to professional development programs. Of course, it is not possible to know if the professional development sources were supplementary to initial teacher education or were the first opportunity teachers had to develop competence in assessment. Nonetheless, few participants rated themselves as lacking competence.

Table 2. Teacher self-rated competence for assessment duties.

Response option	Primary		Secondary		Total	
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
I feel I lack competence for assessment	2	1.7	29	6.5	31	5.5
I feel competent, thanks to my initial teacher training	33	28.4	117	26.1	150	26.5
I feel competent thanks to professional development programs I attended.	67	57.8	255	56.9	322	56.9
I feel competent thanks to my own self-learning initiative.	14	12.0	47	10.5	61	10.8
Subtotal	116		448		564	

Instruments

Remesal's Model (QMCoA). From Remesal's (2011) study a self-report questionnaire of 40 items was designed (QMCoA-Qualitative Model of Conceptions of Assessment). This inventory was tested on a world-wide sample of Spanish as a Foreign Language teachers working in different teaching contexts (Remesal & Brown, 2015). The inventory uses a bifactor approach in which each item is treated as a manifest variable of two factors. One factor has to do with the domain (i.e., learning, teaching, accounting, or certifying) and the second aspect consists of either a formative-regulatory view or a non-regulatory accreditative tendency. The current study used the Spanish language version of the QMCoA.

Brown's Model (TCoA). Brown's Teachers' Conception of Assessment (TCoA) questionnaire consists of 27 items in a hierarchical structure (Brown, 2006). The abridged inventory has nine factors, each with three items, and seven factors are nested under two superordinate factors. The inventory is thus structured around four purposes underlying assessment: (a) improvement of teaching and learning, (b) demonstrating school accountability, (c) certifying student accountability, and (d) treating assessment as irrelevant. The inter-correlation of these four factors combined with their mean scores constitute

teachers' conceptions of assessment. The Spanish adaptation and translation of this inventory (Brown & Remesal, 2012) developed in Spain was used in this study.

Response Scale. Both the QMCoA and TCoA questionnaires used a positively packed agreement rating scale to enhance comparability. Positively-packed scales have more positive options than negative options (Brown, 2004; Lam & Klockars, 1982) on the assumption that participants will tend to agree, a valid assumption when probing the opinions of teachers employed to implement a jurisdiction's policies. This approach also increases the variance of responses, leading to more accurate psychometric properties. Although the scale is a six-point, ordinal agreement scale, maximum likelihood estimation with Pearson product moments was used since scales of this length can be treated as continuous (Finney & DiStefano, 2006).

Analysis

Data Preparation. All cases with more than 10% missing values within each inventory were dropped from analysis and missing values in the balance of data were imputed using the expectation maximization procedure (Dempster, Laird, & Rubin, 1977). Three participants failed to meet this standard for at least one inventory, resulting in a final sample of 187 teachers from 2014 and 376 in 2017.

Model Development & Testing. The analytic strategy was to establish a well-fitting measurement model for each inventory prior to building a structural model that identifies paths between inventories. Hence, the current sample's responses to each inventory was tested against the previously validated factor models, using confirmatory factor analysis (CFA). CFA provides a robust approach to determining if a set of pathways within and among factors correspond to the source data by utilising the factor patterns, covariance patterns, and residual or error values within a data matrix (Byrne, 2001). In CFA, relationships between variables and latent factors that are not expected are set to zero, while

the expected relationships are free to load onto their appropriate factors (Byrne, 2001). Large samples, usually >500, are required to provide stable parameter estimates (Chou & Bentler, 1995).

Thus, pre-existing models may not fit new data because (a) individual items might not have simple structure meaning they belong to multiple or different factors, (b) items may not have statistically significant or meaningful loadings on intended factors, (c) model structures, such as hierarchies or paths between factors, cannot be replicated, and (d) factors may not have independence from other factors in the model (Boomsma & Hoogland, 2001; Marsh, Hau, Balla, & Grayson, 1998). These phenomena can occur because of cultural, policy, and/or social differences between the jurisdiction in which the model was developed and the one in which it is being tested (Brown, Harris, O'Quin, & Lane, 2015). Such differences may thus reflect the impact of the environment rather than deficiency of the instrumentation.

Where changes were necessary, the goal was to maximise similarity to the intended model.

Once acceptable measurement models are established, structural equation modelling (SEM) is used to determine statistically significant paths between inventories. Because the data were collected, in each round, within a single session, the relationship between constructs was investigated as correlational (i.e., TCoA factors correlated with QMCoA factors).

Model Fit. In line with current practice (Fan & Sivo, 2007; Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004), a multi-criteria approach for acceptable model fit was adopted; models were not rejected if $\hat{\gamma} \geq .90$, root mean square errors of approximation (RMSEA) and standardized root mean residuals (SRMR) $\leq .08$, and χ^2/df ratio was statistically non-significant ($p > .01$). Models that met these criteria were not rejected. All analyses were carried out in AMOS (IBM, 2013) using maximum likelihood estimation and Pearson product moment correlations.

Results

Measurement Models

Model trimming & fit. After removing five items due to low fit (i.e., e04, ap06, ac07, r02, and r03), the fit of the data to Remesal's QMCoA was acceptable (Table 3). The hierarchical structure of the TCoA was inadmissible due to negative error variances and positive not definite covariance matrix. Removal of three first-order factors beneath Improvement (except Teaching), merging of the two accountability factors, and deconstructing the Irrelevance factor (i.e., Caution separate and Inaccurate and Ignore joined into one factor) produced an admissible model. Inspection of modification indices suggested the Teaching factor should be predicted by both Improvement and Caution and that the irrelevance factor should also point to the three Student Accountability items and one item (ti1: Assessment is integrated with teaching practice) in Teaching and one item (rel2: Assessment results are consistent) in Improvement. Two items (ig1: assessment forces teachers to teach in a way against their beliefs, ir4: assessment interferes with teaching) were subsequently removed for being weakly explained by their respective factors. This resulted in an acceptably fitting model consisting of 25 items organised in four factors and one subordinate factor (Table 3).

Table 3. Fit statistics for separate and joint models

Model	<i>k</i>	χ^2	<i>df</i>	χ^2/df (<i>p</i>)	CFI	Gamma	RMSEA (90% CI)	SRMR
Remesal QMCoA	35	1528.287	695	2.20 (.14)	.83	.95	.044 (.043- .049)	.055
Brown TCoA revised	25	994.23	262	3.80 (.05)	.83	.90	.071 (.066- .075)	.059
Joint Remesal- Brown	60	3509.39	1641	2.14 (.14)	.80	.90	.045 (.043- .047)	.067

Note. *N*=563; QMCoA=Qualitative Model of Conceptions of Assessment; TCoA=Teacher Conceptions of Assessment; *k*=number of items; CFI=comparative fit index; RMSEA=root mean square error of approximation; SRMR=standardised root mean residual

Measurement Model Descriptives. Within the QMCoA, the formative-regulation factor was moderately inverse to the societal-control factor, while the four domain factors were positively inter-correlated (range: $r=.40-.82$). Mean scores for the QMCoA scales were moderately positive (i.e., $4.00 < M < 5.00$) for the four domains with a small difference between strongest and weakest level of agreement ($d=.35$). In contrast the Societal Control factor had a much stronger mean than the Formative Assessment purpose (i.e., $d= 2.38$) (Table 4). Note that previously surveyed Spanish as a Foreign Language teachers had means for all domains that were lower and most strongly endorsed Formative regulation over Societal Control (Remesal & Brown, 2015).

Correlations between factors within the TCoA were highly variable. Accountability and Improvement were almost identical ($r=.90$) reflecting previous results found in China and Hong Kong which also had similar strong associations between improvement and accountability functions. Like New Zealand studies, the correlation between Irrelevance and Accountability was statistically not significant. Otherwise, inter-correlations were weak ($-.18$ to $.39$). Mean scores for the TCoA scales were weakly to moderately positive (i.e., $3.00 < M < 4.00$) with Irrelevance having the lowest mean and Caution the highest ($d=1.57$) (Table 4).

Table 4. Scale and Inter-Battery Inter-Correlations, Means, and Reliability Estimates

Inventory & Scales	# items	TCoA				QMCoA						<i>M</i>	<i>SD</i>
		I	II	III	IV	V	VI	VII	VII	IX	X		
<i>Teacher Conceptions of Assessment (TCoA)</i>													
I. Improvement	12	(.85)										4.63	0.71
II. Irrelevance	9	-0.18	(.57)									3.71	0.69
III. Caution	6	0.26	0.39	(.56)								4.76	0.65
IV. Accountability	6	0.90	0.07	0.27	(.72)							4.15	0.89
<i>Qualitative Model of Conceptions of Assessment (QMCoA)</i>													
V. Teaching	9	<i>0.60</i>	<i>-0.18</i>	—	<i>0.53</i>	(.57)						4.10	0.65
VI. Learning	9	<i>0.14</i>	—	<i>0.11</i>	—	0.58	(.51)					4.24	0.62
VII. Certification	9	<i>0.47</i>	—	—	<i>0.45</i>	0.80	0.50	(.57)				4.22	0.68
VIII. Accounting	8	<i>0.52</i>	—	—	<i>0.58</i>	0.81	0.40	0.82	(.66)			4.34	0.73
IX. Formative Regulation	18	—	<i>0.69</i>	—	<i>0.11</i>	—	—	—	—	(.83)		3.37	0.83
X. Societal Control	17	<i>0.14</i>	<i>-0.09</i>	<i>0.53</i>	<i>0.10</i>	—	—	—	—	-0.39	(.74)	5.01	0.51

Note. Total items for TCoA is 25 but the count sums to 33 because Caution and Improvement both load on sub-factor Improve Teaching which has 3 items. Irrelevance loads on 3 shared items with Accountability and 1 item in Improve Teaching and 1 item in Improvement. Alpha estimate of scale reliability in brackets on diagonal. Within inventory values shown in bold. Inter-battery correlations in italics. Inter-battery values not reported are not statistically significant.

Structural Model: Inter-correlation of QMCoA and TCoA

After removing statistically not significant inter-correlations between the two inventories, acceptable fit for the full sample was found (Table 3). The remaining inter-correlations between the two inventories can be seen in Table 4 while Figure 1 illustrates the stronger relationships between the TCoA and QMCoA factors by hiding the items and inter-correlation values $<.30$.

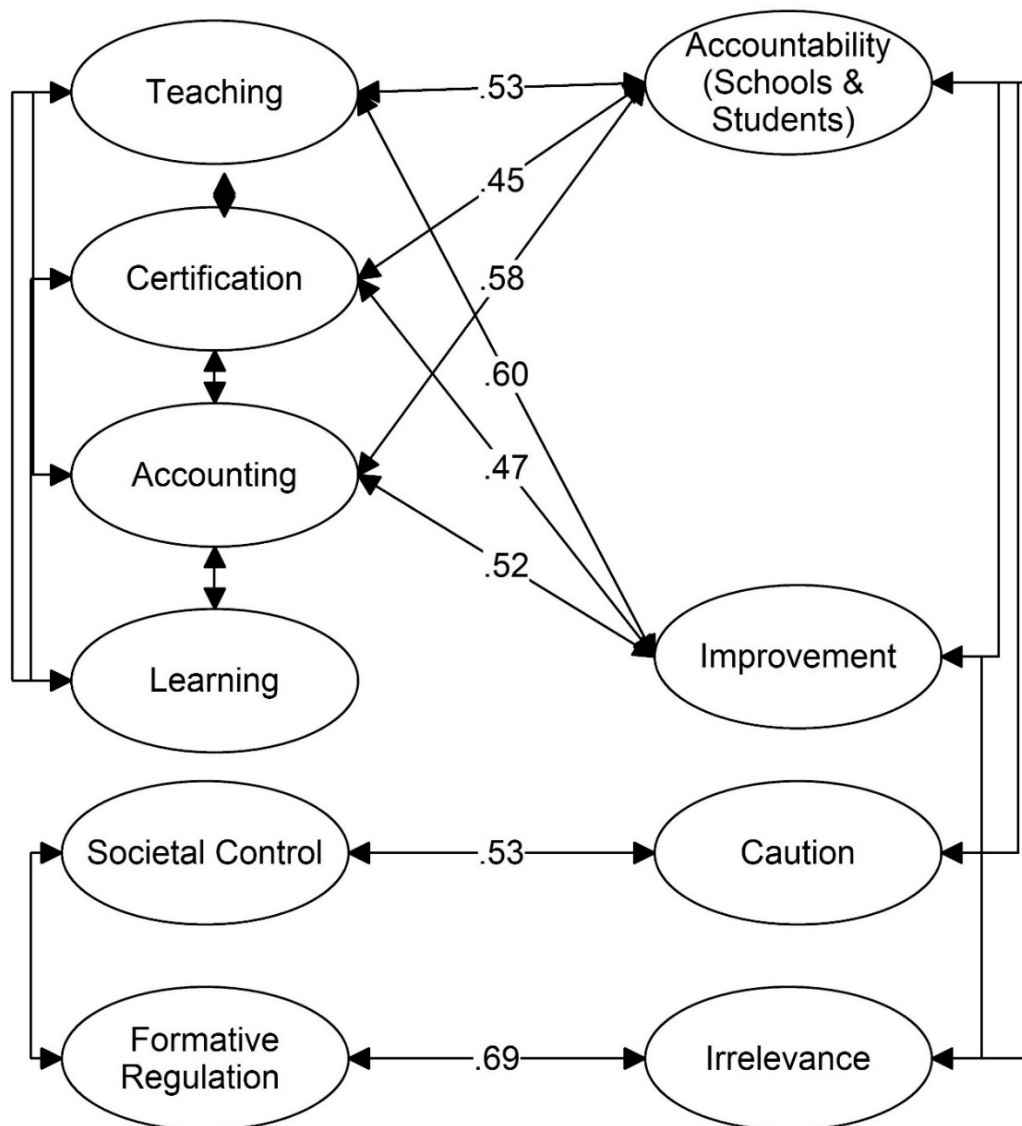


Figure 1. Schematic Model of QMCoA and TCoA factor inter-correlations

Note. Only values $r > .30$ shown for simplicity

Interestingly, both Improvement and Caution have very similar loadings on the QMCoA factors despite having only a weak inter-correlation ($r=.26$). Both factors have moderate correlations ($.45 < r < .60$) on Teaching, Certification, and Accounting (mean $r=.53$). Irrelevance was moderately correlated to just Formative Regulation, whereas Societal Control was moderately correlated with Caution. These paths understandably link assessment for school and student accountability purposes with teaching, accounting, and certifying domains, which focus on the certification of student learning and school success through students' examination performance. Improvement-oriented purposes for assessment are likewise linked to teaching, accounting, and certifying domains. Since accountability and improvement purposes are so highly linked ($r=.90$) it is not surprising that these two purposes correlate with the same domains.

More surprising is positive association of Formative Regulation with Irrelevance, suggesting that efforts to focus on using assessment to improve learning and teaching, as opposed to preparing students for certifying examination, is irrelevant to the main function of school-based assessment. Less surprising is the positive association of Societal Control with Caution, which contains statements focused on treating results cautiously because of measurement error (ir2), taking into account the error and imprecision in all assessment (ir3), or assessment forcing teachers to teach against their beliefs (ir1). This association suggests that teachers are aware of the validity limitations of the examination system and the need to interpret scores appropriately.

Discussion

The aims of this study were, first, to ascertain the conceptions of assessment of this sample of Ecuadorian teachers and, second, to discover the relationship between the two known inventories of teachers' conceptions of assessment (i.e., Brown's TCoA and Remesal's QMCoA). To address this, first the validity of the original models was evaluated

and then the inter-relationship between the inventories was examined. The conceptions of the teachers were examined in three ways: (1) Remesal's QMCoA, (2) Brown's TCoA, and (3) the joint relationship of the TCoA with the QMCoA.

The results showed that Remesal's model had good fit in this large sample of Ecuadorian teachers, after removing five items. A modified version of Brown's TCoA had acceptable fit and the modifications suggest that many items still retained their original meaning. For example all school and student accountability items stayed together, albeit in a merged factor and all improvement items stayed together under the original superordinate factor. Four of the nine Irrelevance items stayed together, while three others joined with the teaching improvement factor to create a Caution factor.

The inter-correlated model of the full QMCoA bifactor structure and the adapted four factor TCoA paints an interesting picture of how this sample of teachers conceived of assessment. The picture is one in which the examination system dominates how assessment is understood. Teaching students for improved performance on accountability examinations was fundamentally a single construct that correlated moderately with teaching, certifying, and accounting. This picture resembles results from Hong Kong (Brown, Kennedy, Fok, Chan, & Yu, 2009), China (Brown, Hui, Yu, & Kennedy, 2011), and Egypt (Gebriel & Brown, 2013) in which powerful high-stakes examination systems create a conception that teaching students for accountability examinations is the best way to improve student learning. Those studies all took place in school systems that are strongly driven by external public examinations with substantial consequences for students and schools. In those contexts, teachers appear to meet their professional obligations by maximising coverage of the tested curriculum and ensuring students receive accurate information about their performance relative to the demands of the examination system.

Unfortunately, this also means that the ‘soft’ policy option (Kennedy, Chan, & Fok, 2011) of assessment for learning (i.e., using assessment to formatively improve teaching and learning with a lower emphasis on summative evaluation) is seen as irrelevant in the presence of the ‘hard’ policy of high-stakes examinations. Consequently, perhaps in response to the power of examinations to control education, teachers associate that fact with the need to exercise caution when interpreting examination scores or results and when letting the examination system control their teaching practices. Perhaps, the teachers grasp that there is more to schooling than teaching to the test and that good teaching is more than teaching to an examination-focused curriculum. Hence, the current results seem to indicate that teachers resolve the call to be formative in their assessment practices by focusing on maximising examination outcomes for all students. This type of tension between accountability-oriented evaluation and improvement-oriented assessment is well-established in recent reviews of teacher perceptions of assessment (Bonner, 2016; Barnes, Fives, & Dacey, 2015). Thus, we conclude that the Ecuadorian teachers experience similar concerns to teachers world-wide.

The results also suggests that, rather than focusing on making teachers believe more strongly than they currently do in improvement-oriented assessment or assessment for learning, efforts need to be made to reduce the dominance of the public examination system with its high-stakes. This argument, arising from the low-stakes assessment environment of New Zealand, begins with the assumption that teachers are generally positively oriented towards using assessment to improve their teaching practices and aiding students to improve. However, the system policy environment (i.e., ranking of students and schools through examination scores and awarding opportunities through examination scores) in which teachers work requires them to focus on high-stakes accountability-oriented practices and priorities. Thus, it has been argued (Brown, 2004; Brown & Hattie, 2012; Hattie & Brown, 2008, 2010), that unless those constraining conditions are modified to be more supportive of

assessment for learning, teachers' current beliefs in assessment for learning will have little room to be effective. Other systems have successfully lowered consequences attached to a sole terminal examination by allowing a mixed system of certification in which school-based, teacher-made evaluations contribute to overall qualifications decisions (Crooks, 2010). Alternatively, greater investment in opportunities for further education (i.e., more funded spaces in universities, more polytechnic places, or apprenticeship programs) and/or the development of more pathways to gainful employment (e.g., second chance vocational educational systems) would reduce the negative consequences of doing poorly or even doing just acceptably on public examinations.

The differences in these results seem consistent with the cultural and socio-economic conditions of Ecuador. While the analyses are plausible, the survey is not conclusive; rather, it would be more persuasive to examine the impact of a natural experiment in which a major policy shift is introduced into Ecuadorian education or if a large enough group of teachers from Ecuador could be tracked as they engaged in education in a different educational system (or vice versa). While some examination-driven societies have introduced 'assessment for learning' policies (Berry, 2011), it needs to be kept in mind that such an introduction needs to be a serious policy change (i.e., minimising the control of the examination system), not simply a matter of adding a new 'soft' policy to an existing 'hard' policy (Kennedy, Chan, & Fok, 2008).

The study can be criticised as being exploratory because it depends on a convenience rather than representative sample and because no previous work with these instruments in this context had taken place. Hence, it is possible that the current results are entirely a function of artefacts of chance factors within the sampled participants. For example, the sample is largely middle-aged and experienced teachers studying towards a master's degree; this suggests that the results may have little generalisability toward new young teachers or prospective teachers

in teacher education. Furthermore, the analytic procedure to establish a well-fitting model for the TCoA inventory does mean that the results may be a consequence of the data-driven model trimming decisions taken. That there is conceptual interpretability of the revised models does suggest the decisions have some warrant. International studies with the TCoA tend to replicate item aggregation into the original factors which have quite different relations to each other; these differences are normally interpreted as reflecting the impact of the local ecology (e.g., Brown & Michaelides, 2012). Nevertheless, although the sample is sufficiently large to allow structural equation modeling to be undertaken, future studies elsewhere in Ibero-America may generate different results and test the current results. Additionally, given that the two inventories were developed in educational contexts (i.e., strong emphasis on low-stakes assessment practices and policy) that differ from Ecuador's high-stakes examination framework; future studies might develop constructs and items that more closely reflect the circumstances and tensions present in that country. Thus, we offer the current report as the beginning of what we hope will be more extensive research into teacher conceptions of assessment in South America.

Nonetheless, policy makers, teacher development and teacher education officials, and assessment experts in Ecuador can take heart that teachers in their country have a positive view towards an educational use of assessment. What is needed now are changes to the system to allow greater space for formative approaches to assessment and schooling in general. Such changes, unsurprisingly, are difficult since extensive time-consuming collaboration with parents and teachers is needed to reach a positive use of assessment for learning.

References

- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. Gregoire Gill (Eds.), *International Handbook of Research on Teacher Beliefs* (pp. 284-300). New York: Routledge.
- Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Policy, Principles and Practice*, 18(2), 199-211.
- Bonner, S. M. (2016). Teachers' Perceptions about Assessment: Competing Narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 21-39). New York: Routledge.
- Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. Du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 139-168). Lincolnwood, IL: Scientific Software International.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11(3), 301-318. doi: 10.1080/0969594042000304609
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166-170. doi: 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: Nova Science Publishers.
- Brown, G. T. L., & Hattie, J. A. (2012). The benefits of regular standardized assessment in childhood education: Guiding improved instruction and learning. In S. Suggate & E. Reese (Eds.) *Contemporary Educational Debates in Childhood Education and Development* (pp. 287-292). London: Routledge.

- Brown, G. T. L., & Michaelides, M. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education, 26*(3), 319-337. doi: 10.1007/s10212-010-0052-3
- Brown, G. T. L., Harris, L. R., O'Quin, C., & Lane, K. E. (2015). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *International Journal of Research & Method in Education, 1*-25. doi:10.1080/1743727X.2015.1070823
- 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. doi: 10.1016/j.ijer.2011.10.003
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347-363. doi:10.1080/09695940903319737
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*(1), 210-220. doi: 10.1016/j.tate.2010.08.003
- 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. 10.5209/rev_SJOP.2012.v15.n1.37286
- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: LEA.

- Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37-55). Thousand Oaks, CA: Sage.
- Crooks, T. J. (2010). Classroom assessment in policy context (New Zealand). In B. McGraw, P. Peterson, & E. L. Baker (Eds.), *The International Encyclopedia of Education* (3rd ed., pp. 443-448). Oxford, UK: Elsevier.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39(1), 1-38.
- Dussel, I. (2001). La formación de docentes para la educación secundaria en América Latina: perspectivas comparadas. *Los formadores de jóvenes en América Latina. Desafíos, Experiencias y Propuestas. Oficina Internacional de Educación*, 10-23.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509–529.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. D. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age Publishing.
- Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy, and Practice*. doi:10.1080/0969594X.2015.1017445
- Gebril, A., & Brown, G. T. L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, 21(1), 16-33.
doi:10.1080/0969594X.2013.831030

- Hattie, J. A. C., & Brown, G. T. L. (2008). Technology for school-based assessment and assessment for learning: Development principles from New Zealand. *Journal of Educational Technology Systems, 36*(2), 189-201. doi:10.2190/ET.36.2.g
- Hattie, J. A., & Brown, G. T. L. (2010). Assessment and evaluation. In C. Rubie-Davies (Ed.), *Educational Psychology: Concepts, Research and Challenges* (pp. 102-117). Abingdon, UK: Routledge.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
- IBM. (2013). Amos [computer program] (Version 23, Build 1812). Wexford, PA: Amos Development Corporation.
- Kennedy, K. J., Chan, J. K. S., & Fok, P. K. (2011). Holding policy-makers to account: exploring 'soft' and 'hard' policy and the implications for curriculum reform. *London Review of Education, 9*(1), 41-54. doi:10.1080/14748460.2011.550433
- Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement, 19*(4), 317-322.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341.
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181-220.
- Ministerio de Educación (2016). *Rendición de cuentas 2016. Ministerio de Educación*. Ecuador.

- Nichols, S. L., & Harris, L. R. (2016). Accountability assessment's effects on teachers and schools. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 40-56). New York: Routledge.
- OECD. (2016). *Making Education Count for Development: Data Collection and Availability in Six PISA for Development Countries*. Paris, FR: OECD Publishing.
- Ponce, J., Bedi, A. & Vos, R. (2002) *¿Cómo hacer más eficiente el gasto educativo? Un análisis de los determinantes del acceso a la educación básica y secundaria en el Ecuador*, SIISE. Quito.
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: the Spanish instance. Building upon Black and Wiliam (2005). *The Curriculum Journal*, 18(1), 27-38.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472-482.
- Remesal, A., & Brown, G. T. L. (2015). Conceptions of assessment when the teaching context and learner population matter: Compulsory school versus non-compulsory adult education contexts. *European Journal of Psychology of Education*, 30(3), 331-347. 10.1007/s10212-014-0236-3
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39-83). Chicago, IL: Rand McNally.
- Torres, R.M., (2006). Analfabetismo y alfabetización en el Ecuador. Opciones para la política y la práctica. Paper presented at the *Education for All Global Monitoring Report 2006. Literacy for Life*
- Viteri, G. (2006). *Situación de la educación en Ecuador*. Gov. Ecuador. Ministerio de Educación.

Walsh, C. (1998). La interculturalidad y la educación básica ecuatoriana: Propuestas para la reforma educativa. *ProcesoS, Revista Ecuatoriana de Historia*, 12, 119-128.