

Brown, Gavin T L; Remesal, Ana  
Prospective Teachers' Conceptions of Assessment: A Cross-Cultural Comparison  
The Spanish Journal of Psychology, vol. 15, núm. 1, 2012, pp. 75-89  
Universidad Complutense de Madrid  
Madrid, España

Available in: <http://www.redalyc.org/articulo.oa?id=17223141008>



*The Spanish Journal of Psychology*,  
ISSN (Printed Version): 1138-7416  
[psyjour@sis.ucm.es](mailto:psyjour@sis.ucm.es)  
Universidad Complutense de Madrid  
España

## Prospective Teachers' Conceptions of Assessment: A Cross-Cultural Comparison

Gavin T L Brown<sup>1</sup> and Ana Remesal<sup>2</sup>

<sup>1</sup>The University of Auckland (New Zealand)

<sup>2</sup>Universidad de Barcelona (Spain)

This paper examines the responses of two samples of prospective teachers (New Zealand,  $n = 324$ ; and Spain,  $n = 672$ ) to the Teachers' Conceptions of Assessment inventory (English and Spanish respectively). The inventory captures four major intentions for assessment (i.e., improvement, irrelevance, school and student accountability). The conceptions of prospective teachers about the nature and purpose of assessment are relevant, given that (a) much educational assessment is carried out in classrooms; and (b) prospective teachers enter the teacher education programs with significant prior school experience of assessment as pupils. Results of confirmatory factor analysis indicated that the original model was inadmissible and that the best-fitting revised model was only configurally invariant between the two samples. It would appear that lack of teaching experience results in different responses for prospective teachers to those of practicing teachers. Moreover, differences in societal and cultural priorities for assessment use most likely explain the lack of invariance between samples.

*Keywords:* Prospective teachers, conceptions, assessment, multi-group confirmatory factor analysis.

En este trabajo se examinan las respuestas de dos muestras de estudiantes de magisterio (Nueva Zelanda,  $n = 324$ ; y España,  $n = 672$ ) al cuestionario Teachers' Conceptions of Assessment (en sendas versiones en inglés y español). El cuestionario recoge cuatro intenciones principales de la evaluación en la acción docente (mejora de los procesos de enseñanza y aprendizaje, irrelevancia, rendición de cuentas de la institución escolar y rendición de cuentas del alumno). Las concepciones de los estudiantes de magisterio sobre la naturaleza y los propósitos de la evaluación del aprendizaje son relevantes dado que (a) la evaluación es una actividad frecuente en el aula; y (b) los estudiantes de magisterio inician la formación con una amplia experiencia previa como alumnos. Los resultados del análisis factorial confirmatorio indican que el modelo original no es admisible para estas muestras; en consecuencia, se presentan y discuten modelos alternativos. Las diferencias explicativas más plausibles encontradas apuntan hacia diferencias culturales y de experiencia docente.

*Palabras clave:* formación del profesorado, concepciones, evaluación, análisis factorial confirmatorio multi-grupo.

---

Ana Remesal is member of the Development, Interaction and Communication in Educational Contexts Consolidated Research Grup. This group has been supported and financed by the Generalitat de Catalunya since 1995 (2009 SGR 933).

The assistance of Sheryll McIntosh, Senior Tutor in the Liberal Arts program of the Faculty of Education, The University of Auckland is acknowledged.

This paper was written while Gavin Brown was an Associate Professor at the Hong Kong Institute of Education.

Correspondence concerning this article should be addressed to Gavin T L Brown. School of Teaching, Learning & Development, Faculty of Education, The University of Auckland. Private Bag 92019, Auckland, 1142 (New Zealand). E-mail: [gt.brown@auckland.ac.nz](mailto:gt.brown@auckland.ac.nz)

Despite numerous similarities, school systems differ worldwide in terms of the role teachers have concerning the assessment of student learning. In some countries (e.g., USA or UK) teacher assessment practices are largely oriented around the powerful external assessment systems; while, in other countries (e.g., Spain or New Zealand), teachers largely practice low-stakes, classroom-based assessment. There is evidence that teachers' assessment techniques and practices are influenced by their beliefs about the nature and purpose of assessment (Brown, 2009; Coll & Remesal, 2009; Delandshere & Jones, 1999). There is also evidence that teacher belief systems differ from society to society in that teachers' conceptions tend to be consistent with the policy and cultural priorities of any jurisdiction (Brown & Harris, 2009; Brown, Lake, & Matters, 2009, 2011).

The relationship between beliefs or conceptions and school practices has been well established in previous studies (van den Berg, 2002). Teacher's beliefs about assessment may hinder innovative changes (Delandshere & Jones, 1999; Remesal, 2007). Griffiths, Gore, and Ladwig (2006) reported that beliefs affect teaching practices to a greater degree than teaching experience and socioeconomic context do. Thus, as long as teachers implement assessment policies in school contexts, the nature and structure of teachers' beliefs about assessment will matter for how and why assessment is carried out. Further, as long as societies do not have identical priorities or practices, we can expect systematic variation in teacher conceptions.

Wolf, Bixby, Glenn, and Gardner (1991) distinguished between two opposite poles in a continuum; that is, the 'assessment culture' and the 'testing culture'. The different ideas teachers held about (a) intelligence, (b) the process of teaching and learning, (c) the nature of assessment tasks, and (d) evaluation criteria shaped their understanding and practices of assessment. Delandshere and Jones (1999) proposed three dimensions to identify and describe teachers' beliefs about assessment: (a) purposes and functions of assessment, (b) teachers' perception of curriculum and their professional self-efficacy feeling, and (c) their beliefs about the teaching and learning process and about students as learners.

Brown and colleagues have developed a comprehensive research program on teachers' conceptions of assessment since the early 2000's. They grounded their research on Thompson's definition of conceptions as: "a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like" (Thompson, 1992, p. 130) and Ajzen's (2005) model of planned behaviour in which intentions or purposes are powerful predictors of behaviour. In this paper, we focus on Brown's (2008) model of teachers' conceptions of assessment which aggregates teacher thinking about the nature and purpose of assessment into four major purpose-related beliefs (i.e., assessment is for improving teaching

and learning; assessment evaluates and holds schools and teachers accountable; assessment certifies students' learning and holds them accountable; and assessment is irrelevant). Indeed, the improvement, accountability, and irrelevance purposes, when conjoined with a distinction between school and student-focus, appear to capture many of the distinctions teachers make about the uses of assessment (Harris & Brown, 2009). Survey research with teachers in New Zealand and Queensland using the Teachers' Conceptions of Assessment (TCoA) inventory (Brown, 2006, 2008) has shown that teachers tended to identify the improvement of teaching and learning as their dominant purpose for assessment. Further, this purpose was but weakly correlated with the purpose of grading or evaluating students and more negatively correlated with the conception that assessment is bad, unfair, and inaccurate.

Additional survey research with New Zealand secondary students using the Students' Conceptions of Assessment inventory has shown that students also believed that improvement was the dominant purpose for assessment (Brown, Irving, Peterson, & Hirschfeld, 2009). Another study showed that the conception of assessment as evaluating the student through grading was positively associated with increased learning outcomes (Brown & Hirschfeld, 2008). Hence, it appears that when students agreed with being evaluated or graded their grades increase; however, this practice is not usually considered to be a legitimate part of assessment *for* learning practices that teachers are expected to implement (e.g., Leahy, Lyon, Thompson, & Wiliam, 2005; New Zealand Ministry of Education, 2007; Ministerio de Educación Español, 2007a, 2007b, 2007c). Thus, it may be that one of the most important changes that prospective teachers need to go through involves their conceptions, that is, their understanding, beliefs, and attitudes towards the uses of assessment.

Pajares (1992) suggested that the belief systems of teachers and prospective teachers come predominantly from their experiences as students; if this was true, it would be expected that prospective teachers would have similar belief patterns to those of practicing teachers. However, our understanding of the belief systems of prospective teachers is quite weak. Some evidence for this would come from items from the TCoA aggregating into the same factors as they did for practicing teachers. However, it needs to be borne in mind that students and teachers experience school practices from different perspectives with different roles and power. In other words, teachers assess, evaluate, test, and grade students and not the other way round. Hence, it may be that prospective teachers, while they are still in Teacher Education classes, may not have fully transitioned into the responsibilities of teaching practice. For example, prospective teachers might more easily endorse the grading, evaluative purpose of assessment than the formative purpose simply because they have not had experience of being a

teacher using student assessments to modify teaching practice. Therefore, we might expect some differences in conceptions of assessment between prospective and practicing teachers, especially when new assessment policy initiatives have not yet had enough time to impact on prospective teachers' prior experiences of schooling (Bertoret & Artiga, 2004). Evidence for this would come from TCoA items belonging to different factors or there being different structural relations between factors for prospective teachers as compared to practicing teachers.

Furthermore, there is a strong probability that, inasmuch as educational assessment practices are different between societies, teachers in different systems will have differing conceptions of assessment. Cross-cultural research with the TCoA has suggested that the strong historical and cultural acceptance of the importance and value of public examinations in Chinese contexts impacts on teacher beliefs systems. For example, Hong Kong teachers strongly associated assessment for improvement with assessment as making students accountable (Brown, Kennedy, Fok, Chan, & Yu, 2009). Brown and Harris (2009) argued that recent changes in the way a formative assessment tool was used in New Zealand schools for school-wide, evidence-based improvement initiatives led to a significant shift in New Zealand Teacher's conceptions of assessment; that is, they had a much stronger emphasis on school accountability than previous studies had found. It seems likely that social and cultural priorities express themselves in policies and practices and that these mutually and interactively influence teacher's beliefs and are shaped by the beliefs of individuals in a society (Bandura, 2001). Hence, it is likely that to the degree countries have different educational assessment systems and policies and different cultural priorities, teachers would have different beliefs. Whereas both Spain and New Zealand have adopted an assessment for learning policy, there are critical differences in implementation and context for the policy, the most important of which is probably related to the much longer absence of high-stakes testing in primary schools in New Zealand (details below). Hence, while it seems possible that prospective teachers in the two societies would have similar conceptions of assessment because of the similar policy framework, the cultural, linguistic, and historic differences between societies may be stronger than the policy similarities, which are much more recent.

### *Assessment Contexts*

*Spain.* Spain has gone through important educational systemic changes since the 1990s. Previously, compulsory basic education lasted eight years and after the 8<sup>th</sup> grade, non-compulsory education followed. Secondary schooling was either four years of high school leading to college and university, or alternatively, three to five years of basic or middle vocational training leading to employment. The

LOGSE (*Ley de Ordenación General del Sistema Educativo*, 1990) re-structured the school system: compulsory education was extended to ten years in line with the majority of European school systems. In addition, these ten years of schooling were redistributed in a '6 + 4' scheme (i.e., six years of Primary School and four years of Secondary School). Three options exist upon completion of compulsory secondary education: that is, (1) a basic vocational education, (2) a technical vocational education, or (3) two more years of high school leading to university or superior vocational education. While, further small reforms have been promulgated, coincident with changes of national government, none has affected this main structure of Spanish schooling (Remesal, 2007).

The curriculum is progressively defined at three levels: State Basic Common Curriculum, Autonomous Community Curriculum, and School Curriculum. General outcomes and objectives are defined but each school is expected to adjust and redefine the basic lines to its own immediate social context. Assessment is low-stakes, school-based, meant to be continuous, formative, and holistic. Promotion decisions, both at primary and compulsory secondary education, are based on a majority agreement of the whole school teaching staff. No external evaluation whatsoever is done on a regular basis at the national level; even the Graduate of Compulsory Secondary Education (*Grado de Educación Secundaria Obligatoria*) certificate is obtained after a school-based assessment process that takes into account the students' holistic learning progress from grades 7 to 10. However, Catalonia, the autonomous community in which data for this study were collected, is experimenting with initiatives towards external evaluation of pupils' competences in basic curricular areas at the end of primary education. At the end of post-compulsory high-school, a university entrance examination must be passed in order to pursue a higher education career. Hence, more than fifteen years after the 1990 structural reform, we could expect that Spanish compulsory education teachers would be mainly in favour of using assessment for improving teaching and learning, along with the use of assessment to evaluate school quality and, to a lesser extent, to certificate students' learning.

*New Zealand.* In the last two and a half decades large structural changes have been initiated in New Zealand schooling and education (Fiske & Ladd, 2000; Levin, 2001). The national assessment policy in the primary school sector emphasizes voluntary, school-based assessment for the purposes of raising achievement and improving the quality of teaching programmes (New Zealand Ministry of Education, 1994, 2007) relative to the outcomes objectives specified in the national curriculum. The curriculum itself is child centred, non-prescriptive (i.e., there are very few textbook series used in primary schools and no examination syllabi to control teaching), holistic, and integrated (Crooks, 2002) while, simultaneously, having managerial overtones with common nationally-specified outcomes and objectives

across eight levels. There is no compulsory, state mandated assessment regime in the primary school sector (Crooks, 2010); hence, all assessment practices are voluntarily-chosen and controlled at the school level and, consequently, relatively low-stakes. Indeed, even before World War II, end-of-primary selection examinations (i.e., Standards) were removed in favour of teacher-judgements concerning academic progress (Brown, Irving, & Keegan, 2008). However, the national policy has a strong public accountability element in that schools are expected to demonstrate that student performance is improving relative to the curriculum levels and objectives.

Primary school teachers make extensive use of informal assessments and standardised tests (Crooks, 2002), primarily for the purpose of improving instruction and student learning (Croft, Strafford, & Mapa, 2000; Hill, 2000). In contrast to primary school, at secondary school assessment is primarily focused on preparing for or implementing the high-stakes, student qualifications system (i.e., the National Certificate of Educational Achievement) which begins formally in the third year of secondary schooling when students are about age 15 (Crooks, 2010). Hence, we should expect that teachers in New Zealand are strongly committed to the notion of assessment for improved learning and teaching, nor reluctant to use assessments to evaluate student learning, especially for the purpose of informing improved teaching.

### Research Questions

Given the transition that prospective teachers must undergo in their process of becoming teachers, this study evaluated the measurement model of the Teachers' Conceptions of Assessment Abridged (TCoA-IIIa) inventory (Brown, 2001-2003) with New Zealand and Spanish prospective teachers enrolled at the institutions of the two authors. The research questions examined were:

1. What measurement model best describes how prospective teachers from both sites respond to the Teachers' Conceptions of Assessment inventory?
2. Are the responses of the prospective teachers from the two sites invariant?
3. What context variables most likely explain the responses of the prospective teachers from the two sites?

### Method

#### Participants

*Spain.* In the academic year 2008-09, 672 freshmen prospective teachers voluntarily participated in the study. All were taking a one year course in Educational Psychology, which is offered in the first and second semesters of the teacher education program and is delivered

by 10 different instructors following a common syllabus. This syllabus covers Educational and Developmental Psychology but does not include *assessment* as a learning topic. Most (i.e., 80%) of the participants were female; the average age of the sample was 22 years ( $SD = 5$ ). The sample represents 68% of the population enrolled in the course in thirteen different groups divided in special branches (i.e., Infant School,  $n = 148$ ; Primary School,  $n = 113$ ; Foreign Language Education,  $n = 108$ ; Physical Education,  $n = 112$ ; Musical Education,  $n = 91$ ; and Special Education,  $n = 100$ ). The questionnaire was administered between October and December, before the first examination of the course, which took place in January

*New Zealand.* Between 2004 and 2008 (i.e., 2004,  $n = 115$ ; 2005,  $n = 149$ ; 2008,  $n = 60$ ), 324 second-year undergraduate prospective teachers in the same second-year introductory course on classroom assessment were surveyed in the first week of the course as part of the regular classroom instruction. This course, taught by the first author, aimed to deliberately shift prospective teachers towards adopting and implementing an improvement orientation towards assessment (Brown, Irving, & Keegan, 2008). Sex information was available for 261 participants, of whom most were female (i.e.,  $n = 216$ , 83%). Further information about teaching major or level was not available.

#### Instruments

The TCoA-IIIa uses 27 statements, spread equally across nine factors, which are organised into four major, inter-correlated conceptions of assessment (i.e., assessment improves teaching and learning; assessment grades students; assessment shows quality of schools; assessment is irrelevant) (see panel A of Figure 1 for schematic illustration). Two of these major conceptions have sub-factors; improvement had four 1<sup>st</sup> order factors (i.e., improves teaching, improves learning, is valid, describes student learning) and irrelevance had three first-order factors (i.e., is ignored, is bad, is inaccurate). Note that the validity factor consisted of three items traditionally associated with reliability (i.e., consistency, trustworthiness, and dependability of assessment); these terms avoid technical or statistical language and are understood as part of validity because interpretations and actions depend, in part, on the quality of the scores generated by assessment processes (Kane, 2006). Respondents use a six-point, positively-packed, agreement rating scale known to generate discrimination in contexts of social desirability (Brown, 2004a). The TCoA-IIIa measurement model has been found to have acceptable fit for samples of primary and secondary teachers in both Queensland and New Zealand (Brown, 2008).

The TCoA-III was translated into Spanish in two different phases. A first translation, developed by a doctoral researcher for use in Mexico (Esparza, 2007), was evaluated in a preliminary study (Remesal, 2008). This allowed the

identification of several problems: first, the language had several characteristics of the Mexican Spanish variant, which made it difficult for Peninsular Spanish speakers to understand; second, a too literal translation of certain technical jargon also posed some comprehension difficulties for prospective teachers, since the original scale had been designed for practicing teachers. Thus, a revised second translation (Remesal, 2009) was developed paying special attention to: (a) regional dialectical features that concern educational discourse (e.g., in Mexico *classroom* is called *salón de clase*, while in Spain it is *aula*); (b) rewording of certain statistical terms that are uncommon in daily discourse among Spanish prospective teachers, such as *valid* and *reliable*, for the purpose of clarification; and (c) systematic reference to “*evaluación del aprendizaje*” (classroom assessment) as opposed to “*evaluación externa*” or “*evaluación del centro*” (external or school evaluation), due to the polysemic nature of the term *evaluación* in Spain, which is often even confused in the school context with *school term*. The students were alerted to this possible confusion at the time of the questionnaire administration.

### Analysis

A combination of exploratory and confirmatory factor analyses was used to evaluate the TCoA inventory responses. All cases with >3 missing values were dropped from analysis and missing values in the balance of data were imputed using the expectation maximization procedure (Dempster, Laird, & Rubin, 1977). Consequently, 324 New Zealand cases and 672 Spanish cases were available for analysis. Maximum likelihood confirmatory factor analysis of the variance-covariance Pearson correlation matrices, using AMOS software (Arbuckle, 2008), was used throughout to test the validity of the various measurement models for the two samples. When a model is either inadmissible or poorly fitting, alternatives can be taken. Modifications to the model can be taken (e.g., simplifying the model by removing first-order factors) and exploratory factor analysis can be carried out independently on each sample to develop new models. The key requirement is that such modifications are theoretically defensible and, ideally, tested on an independent sample from the same population (MacCallum, 1995; Maruyama, 1998). In this study, four models were developed and cross-validated with the data from a sample drawn from a different population (i.e., Spain or New Zealand).

The first model was the original TCoA model previously validated with New Zealand and Queensland teachers (Figure 1, Panel A). The second model was identical to the original model with a simplified first-order structure (Figure 1, Panel B). Panel C (Figure 1) shows an alternative model guided by exploratory factor analysis on the New Zealand data, while Panel D shows a fourth model developed from exploratory factor analysis with the Spain data. The key features of each model were:

*Panel A* structures the 27 items for conceptions of assessment as four major, inter-correlated factors (i.e., improvement, irrelevance, school accountability, and student accountability), with seven sub-factors for the improvement and irrelevance factors as described earlier. Each factor consists of three items.

*Panel B* adopts the same meta-structure as Panel A, keeping all the 27 items, but reduces the number of sub-factors within improvement and irrelevance by removing 1<sup>st</sup>-order factors that had empirical negative error variances which, because the values were smaller than their standard errors, was assumed to indicate over-factoring.

*Panel C*, derived from an empirical analysis of the New Zealand sample, increases the number of major factors to five but retains three of the same major, inter-correlated factors of the original model in Panel A and keeps all 27 items. The major differences relate to the position of the Valid factor concerning the consistency and dependability of assessments (i.e., moved from Improvement to School Accountability), the splitting of the Irrelevance factor into two factors (i.e., Bad and Ignore which kept the factor Inaccurate as a sub-factor and some shuffling of items within these three factors), and the movement of one item from the Student Accountability factor to the Describe Learning factor under Improvement. Otherwise, four of the nine first-order factors kept exactly the same items as devised in Panel A.

*Panel D*, derived from an empirical analysis of the Spanish sample, retains the original four, inter-correlated factor structure, with sub-factors for Improvement and Irrelevance, but reduces the number of items to 25. The major differences relate to the sub-factor characteristics for Improvement (i.e., the “improves teaching” and “improves learning” gained one item each and the “describes learning” factor disappears) and Irrelevance (i.e., Ignore disappears, Bad increases to five items, and Inaccuracy gains one item). The two items removed from this model were item #20 about determining student qualification standards from Student Accountability and #14 about modification of ongoing teaching from Improvement. Otherwise, only two of the nine first-order factors kept the same items as devised in Panel A.

In line with suggested practice (Cheung & Rensvold, 2002; Fan & Sivo, 2007; Marsh, Hau, & Wen, 2004; Vandenberg & Lance, 2000) models with statistically non-significant  $\chi^2$  per *df*, gamma hat > .90, and root mean square errors of approximation (RMSEA) and standardized root mean residuals (SRMR) < .08 were considered sufficiently close to the data so as to not be rejected. To test invariance of the models between the Spain and New Zealand samples, a multi-group approach with nested model comparisons was undertaken (Byrne, Shavelson, & Muthen, 1989). Once equivalent configuration of paths (i.e., the same solution for free, fixed, and zero paths) is established (i.e., RMSEA < .05), it is possible to progressively test the invariance of

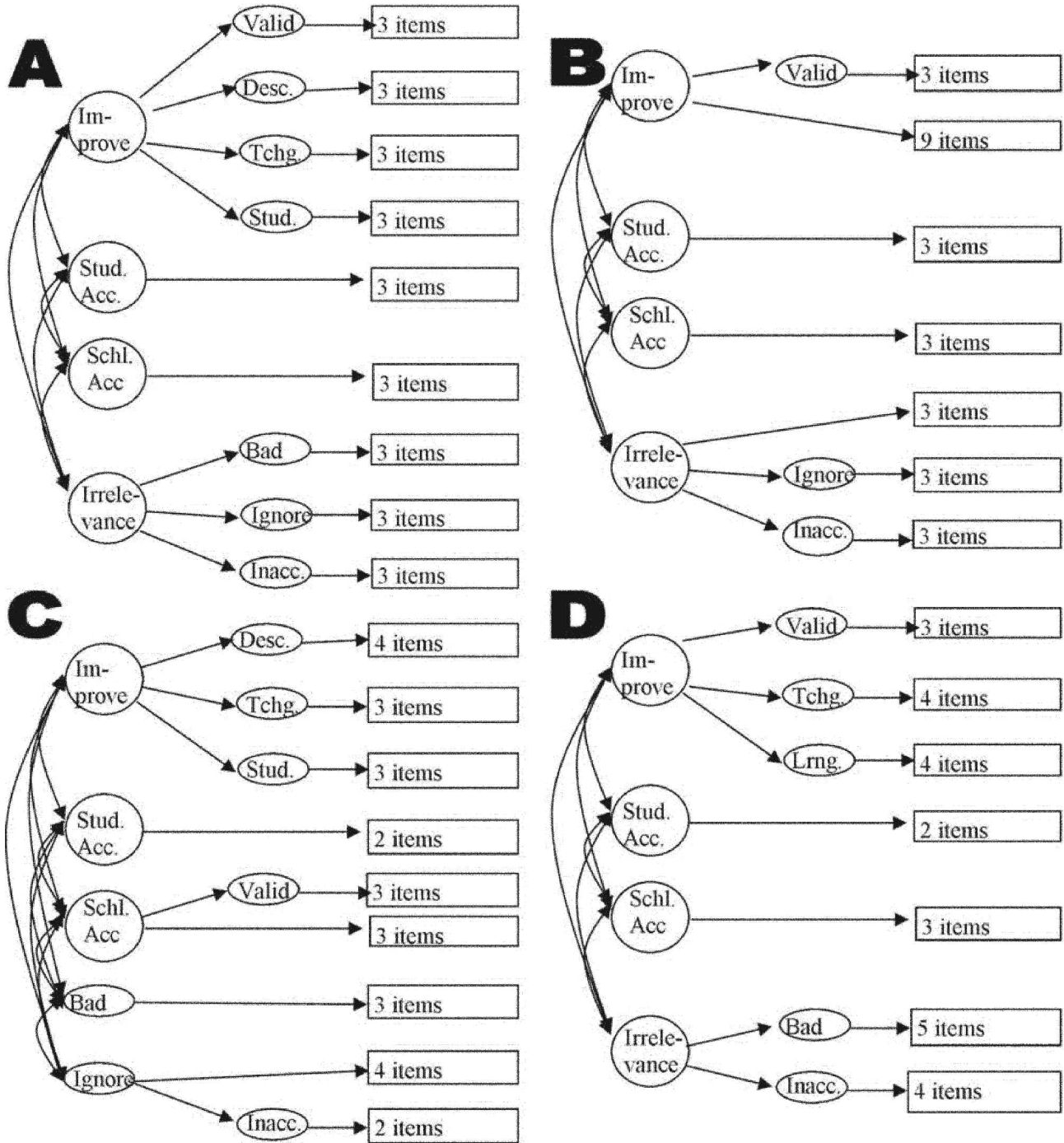


Figure 1. Structure of TCoA models evaluated.

Notes. Panel A = Original NZ model; B = Simplified original NZ model; C = EFA results from NZ undergraduate prospective teachers; D = EFA results from Spain prospective teachers.

the models. Invariance was tested by progressively constraining parameters (i.e., metric invariance of equivalent regression weights, then scalar invariance of equivalent regression intercepts, and then finally equivalent residuals) in each group to be equal to the other group and examining the difference in fit. If the model comparative fit index

(CFI) statistic does not change by more than .01 at each progressively more constrained level, then invariance for that parameter is imputed (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Invariance at the configural, metric, and scalar levels is required to infer the model fits identically in both samples.

Results

Fit statistics based on multi-group confirmatory factor analysis for the four models are in Table 1. The original New Zealand TCoA model was found to be inadmissible due to negative error variances in the first-order factors. Since sample size was > 300 for each group, it is most likely that over-factoring rather than small sample size explains why participants did not make the distinction between these factors that had been previously reported. All three alternate models, as described above, had acceptable fit statistics. Model C, based on an exploratory factor analysis of the New Zealand data, had the best fit characteristics across all indices. Hence, the model in Panel C was used as the basis for evaluating differences in responses to the TCoA between prospective students in the two samples. All three admissible models were tested for invariance between the two groups and only configural invariance was demonstrated (i.e., the  $\Delta CFI > .01$  for equivalent regression weights or metric invariance). Consequently, the current TCoA inventory does not elicit consistent responding between the two samples and the two samples cannot be considered as members of the same population of prospective teachers. Differences in responding suggest clearly there are non-chance differences between

these two samples of New Zealand and Spanish prospective teachers in their conceptions of assessment.

It is worth noting that the estimate of scale reliability for two of the scales in Panel C for the Spanish sample was extremely low (Table 2); nonetheless, the quality of model fit warrants use of the model (Panel C) to examine the belief systems of both groups of prospective teachers. When comparing the two samples, the mean scores for the scales (Table 2) had considerably greater range (i.e., higher maximum and lower minimum) for the New Zealand sample compared to the Spanish sample. In other words, except for the *school quality* factor, Spanish prospective teachers gave very similar mean scores (i.e., between 3.23 and 3.56) to most factors, whereas the New Zealand prospective teachers made strong distinctions between scales they rejected (i.e., *assessment is bad*), those they slightly endorsed (i.e., *assessment is ignored and inaccurate* and *school quality*), and those they moderately endorsed (i.e., *improvement* and *grades students*). Further, the difference in mean scores between the two samples was large (Cohen's effect size  $|d| > .60$ ) for four of the five scales. Only the *assessment is ignored and inaccurate* scale had similar mean scores. The New Zealand sample agreed much more that assessment is for *improvement*, that it measures *school quality*, and that it *grades students*, while the Spanish sample agreed much more that *assessment is bad*.

Table 1  
Fit Statistics and Invariance Results for Four TCoA-TES Models

Model	k	Chi-square fit indices	Goodness of fit indices	Badness of fit indices	Invariance Result	
A. NZ original	54	Inadmissible for both groups due to negative error variance				
B. Simplified NZ original	54	$\chi^2 = 1976.51$ ; $df = 632$ ; $\chi^2/df = 3.127$ , $p = .08$	CFI = .79; GAMMA HAT = .95	RMSEA = .046; SRMR = .068	Configural	
C. EFA NZ*	54	$\chi^2 = 1722.85$ ; $df = 618$ ; $\chi^2/df = 2.788$ , $p = .09$	CFI = .83; GAMMA HAT = .96	RMSEA = .042; SRMR = .062	Configural	
D. EFA Spain	50	$\chi^2 = 1692.33$ ; $df = 529$ ; $\chi^2/df = 3.199$ , $p = .07$	CFI = .79; GAMMA HAT = .96	RMSEA = .047; SRMR = .068	Configural	

Notes. k = number of items; \* = preferred model; Spain N = 672; New Zealand N = 324.

Table 2  
Characteristics of Model C for New Zealand and Spanish prospective teachers (TES)

TCoA-TES Scales	k	New Zealand			Spain			Effect (Cohen's d)
		M	SD	Cronbach $\alpha$	M	SD	Cronbach $\alpha$	
Assessment improves student learning and teaching	10	4.49	.75	.83	3.49	.71	.79	1.38
Assessment is ignored and is inaccurate	6	3.33	.79	.69	3.56	.71	.56	-.31
Assessment is bad	3	2.42	.94	.62	3.23	.79	.29	-.96
Assessment measures school quality validly	6	3.53	.85	.77	2.77	.69	.67	.78
Assessment grades students	2	4.43	.94	.55	3.31	1.06	.18	1.10

Notes. k=number of items; Spain N = 672; New Zealand N = 324.



Examination of the factor inter-correlations also showed large differences in how the samples conceived of assessment (Table 3). The differences in correlation values were all statistically significant, except for the correlation between *assessment is bad* and *assessment grades students*, which was identical for both groups. For the New Zealand prospective teachers, the improvement factor was negatively correlated (although weakly) with the irrelevance factors, and moderately correlated with the accountability factors. In contrast, for the Spanish prospective teachers, there was a moderately strong negative correlation with the irrelevance factors and a strong positive correlation with the accountability factors. The correlation between the two irrelevance factors was much stronger for the Spanish prospective teachers than the New Zealand counterparts. The school quality factor was moderately but negatively correlated with the two irrelevance factors for the Spanish prospective teachers, while it was very weakly correlated among the New Zealand prospective teachers. The pattern of inter-correlations for the student grading factor were similar for both groups, except for its relationship to assessment being ignored and inaccurate which was moderately positive among New Zealand prospective teachers and weakly negative among Spanish prospective teachers. Hence, the New Zealand prospective teachers tended to give higher mean scores but have weaker factor inter-correlations than the Spanish prospective teachers.

### Discussion

This study aimed to discover whether the original model of responses to the TCoA-III A as developed with New Zealand practicing teachers (Brown, 2004b) and validated with Queensland primary school teachers (Brown, 2006) would be replicated in the responses of prospective teachers in New Zealand and Spain. The original model consisted of four major factors arguably capturing endorsement of four major uses and purposes of assessment (i.e., improvement, irrelevance, school accountability, and student accountability). Our discussion of the study revolves around

three comparisons; that is, differences between prospective and practicing teachers, differences between the two samples of prospective teachers, and evaluations of the current TCoA instrument and future developments needed.

### *Prospective vs. Practicing Teachers*

The results showed clearly that the original model was inadmissible for both groups. Exploratory and confirmatory factor analytic procedures were used independently within each sample to identify a plausible alternative model. The best fitting model for both groups was Model C, consisting of five factors (i.e., Assessment improves student learning and teaching, Assessment is ignored and is inaccurate, Assessment is bad, Assessment measures school quality validly, and Assessment grades students). These five factors can be conceptually aggregated into three big ideas—assessment improves, assessment is negative, and assessment shows the quality of schools and students. It would appear that the best fitting model for prospective teachers indicates that such participants have different conceptions of assessment purposes to those of experienced, practicing teachers.

A significant difference between these two samples of prospective teachers and practicing teachers surveyed in New Zealand and Queensland (Brown, 2008) is that the location of the validity of assessments factor (i.e., their dependability #24, consistency #15, and trustworthiness #6) had a very different location despite being made up of the identical items. Remember that valid interpretations and actions depend on the reliability (i.e., dependability, consistency, and trustworthiness) of assessment scores or grades. For practicing teachers this validity factor contributed to the conception of assessment for improvement (i.e., improvement can happen if assessments are valid), while for prospective teachers it contributed to the conception of school accountability (i.e., schools can be held accountable if the assessments are valid). This different location cannot be attributed to a problem in translation since the model was derived from the New Zealand sample and the model showed configural invariance between the two groups.

Table 3

*Factor inter-correlations of TCoA-III A Model C for New Zealand and Spain samples*

TCoA-III A Scales	Inter-correlations				
	I	II	III	IV	V
I. Assessment improves student learning and teaching	—	-.46	-.51	.91	.60
II. Assessment is ignored and is inaccurate	-.15	—	.81	-.40	-.14
III. Assessment is bad	-.28	.44	—	-.58	-.12 <sup>ns</sup>
IV. Assessment measures school quality validly	.55	.12	.00	—	.55
V. Assessment grades students	.49	.30	-.12 <sup>ns</sup>	.45	—

Note. Values below diagonal are New Zealand; above diagonal in italics are Spain; Spain  $N=672$ ; New Zealand  $N=324$ ; ns=not significant.

It may be concerning for teacher education that prospective teachers do not associate educational improvement with valid, dependable assessments. A strong emphasis on this conception of assessment implies that improvement could take place solely on the basis of assessments that have unknown quality characteristics, such as teacher ratings or observations. Additionally, assessment literacy development of prospective teachers might be hindered if they consider that dependable assessments relate only to school evaluation practices rather than to their own individual classroom practices. Nevertheless, this is a likely and unsurprising result if we consider the fact that the participants, from both New Zealand and Spain, have had little experience or knowledge of value-added school evaluation systems; the absolute value of external assessment results is normally taken as the indicator of school quality (i.e., good schools have high scores on valid assessments).

The second key difference between the prospective and practicing teachers' conceptions of assessment is related to the location of an item designed as part of student accountability (i.e., evaluating against achievement standards #20). For prospective teachers this item became associated with improved student learning. It may be that, because prospective teachers are still students themselves, it is easier for them, relative to practicing teachers, to consider grading as an improvement function rather than an accountability function. The competing meanings of 'improvement' and 'accountability' conceptions of assessment for prospective and practicing teachers are certainly key areas for future comparative studies. Nonetheless, validation of this result with other samples of prospective teachers is needed to ensure that the current result is not an artefact of sampling.

#### *New Zealand vs. Spanish Prospective Teachers*

Multi-group nested invariance testing found that there was only configural equivalence for both samples, meaning that the same paths were admissible for both groups despite their differences in language and society. However, the regression weights and structural covariances differed by more than chance, indicating that the two samples were clearly from separate populations. Examination of the differences in scale reliabilities, scale means, and scale inter-correlations confirmed that the two samples had very different responses to the instrument. The significant differences between the two groups are entirely consistent with the underlying notion that beliefs, opinions, and attitudes towards educational processes are an ecologically rational response to cultural and societal priorities (Brown & Harris, 2009; Brown, Lake, & Matters, 2009, 2011; Remesal, 2007, 2011). In other words, teacher conceptions or beliefs tend to be aligned with the dominant uses and purposes assigned to assessment within each society. Hence, we now consider plausible context variables that might have some influence on the cross-sample differences.

One possible explanation for the differences could be cultural factors in indicating one's level of agreement with these statements (i.e., a response bias). It is clear that the mean scores for these factors are much lower and closer together for the Spanish sample. This might reflect either a reluctance to indicate one's opinion or a reluctance to express extreme opinions.

A second source of differences might be found in the context of assessment in Spain compared to that of New Zealand. A very common practice in Spanish schools is to establish within-school achievement levels for differential grouping of students according to performance in basic curricular areas. In contrast, in New Zealand grouping tends to be done within classes according to teachers' perceptions of student ability and their preferred understanding of optimal classroom arrangements (Wilkinson & Townsend, 2000). The Spanish approach might be understood pedagogically as a quite acceptable practice for attuned teaching and catering to different learning needs. However, it may be seen from the students' view as a measure of discriminating between good students against weak ones. Hence, this well-intentioned pedagogical measure could have indeed negative effects on students. And since prospective teachers were recently students, they may well be sensitive to problems in this use of assessment.

A third possibility is that differences are influenced by an instructional effect. The New Zealand prospective teachers were enrolled in a specific Assessment course, while the Spanish participants were enrolled in an introductory Educational Psychology course without any focus on assessment in schooling contexts. Therefore, it seems plausible, given the effect of conceptual priming, that the New Zealand participants might have reflected more about assessment, so as to become mentally attuned to the topic, even before the course began, while the Spanish participants had more of a naïve "pure student" view.

Another important source of differences for these results could be located in the important demographic differences in the sampling. While the New Zealand sample involved second-year prospective teachers, the Spanish data were captured from first-year prospective teachers. It is impossible to tell from two cross-sectional surveys of very different culture groups whether there is a developmental trajectory in teacher conceptualisation of assessment. However, the differences may be pointing to the possibility that, along the life-span of teacher experience, conceptions of assessment may change. While, this argument confronts Pajares' (1992) assumption that beliefs and conceptions are relatively stable, there is strong evidence that teacher conceptions of assessment are consistent with the uses and purposes of their employment circumstances (Brown & Harris, 2009; Brown, Lake, & Matters, 2011). While extremely speculative given the lack of longitudinal data, the possibility that beliefs are not immutable is an important opportunity for teacher educators seeking to develop

assessment for learning beliefs and practices among prospective teachers.

### *Current vs. Future TCoA Inventories*

The statistically significant differences between samples suggests that, while the current TCoA inventory in its Spanish translation can work in Spain, a new set of items, factors, and structures is needed to capture the full range of beliefs of prospective teachers, and possibly also of practicing teachers, in the Spanish context. Furthermore, the results raise doubts as to the sufficiency of the four-intention model for capturing the complexity of teacher conceptions. Indeed, when combined with other cross-cultural research with Chinese teachers (Brown, Kennedy, et al., 2009; Li & Hui, 2007), it has become apparent that modelling teacher conceptions with just four intentions may be a function of the cultural origins of the research and that in different societies somewhat different intentions and patterns exist.

While the Spanish model (Panel D) was not used in this study, future development work in Spain should consider the significance of the two items rejected by that model (#20, #14). It may be that the construct of standards (#20) did not fit because in Spanish schools the expected standards or results are seldom communicated to students in everyday classroom practices; in other words how would students or prospective teachers know what is required if this is rarely made explicit? Indeed, translating the notion of standards as a technical term proved problematic; the phrase “niveles de competencia o habilidad requeridos” (i.e., required levels of competency or skill) was used as a functionally equivalent expression. However, even this paraphrase might be too ambiguous for first year prospective teachers since they lack the experiential knowledge of how standards are operationalized in Spanish schooling. Concerning item #14, about changing teaching as a consequence of assessment information, research with teachers in Spain (Remesal, 2006) suggests that they don't usually make teaching decisions visible to their students. Hence, students are generally unaware of teachers' pedagogical goals and plans and teachers often make changes in teaching ‘on the run’ without explaining the changes to their students. Instead, teachers make their pedagogical responses to assessment normally visible only to external and ‘superior’ audiences (e.g., administrators, families, and policy makers). Thus, it is not surprising that the prospective teachers in Spain did not endorse the assessment purpose of modifying teaching. Together, these problematic items for Spanish prospective teachers reinforce the notion that their beliefs are partly a function of policies and practices embedded in a society.

Finally, further troubles with the translation of the inventory must be considered. The term *evaluación*

corresponds in Spanish to two different terms in English, namely *external evaluation* as well as *classroom assessment*. In addition, the term *evaluación* is traditionally and frequently misused in the school context to mean *school term*. This comes about because traditionally final summative assessment (i.e., *evaluación*) is carried out at the end of each term (i.e., *evaluación*). Although the participants were alerted to this misuse in the questionnaire instructions, it is possible the common parlance interfered with participant responses. In other words, there is a need for a substantial revision of the original instrument in order to make it suitable for the Spanish context.

Nonetheless, 23 items from the original TCoA-III A in English were kept in the preferred Model C in factors that were logically equivalent to their original design. Given that the model and inventory were developed for use and validated with practicing teacher populations, it seems quite promising to discover that the same items have similar meaning across languages and cultures and have some applicability to participants who are not yet fully-fledged teachers. The results suggest that there are some potentially universal constructs; moreover, they point to directions by which the inventory could be developed further for use with prospective teachers in any society. The commonality appears to be still consistent with Pajares' (1992) argument that teacher beliefs arise from experiences as students. This leads us to consider that teacher beliefs may reflect the many similar functions for assessment in school contexts across cultures, societies, and languages.

Furthermore, the differences reinforce our understanding that teacher beliefs very much reflect societal priorities and jurisdictional practices. As context changes, so does thinking among teachers and among prospective teachers. Thus, beliefs and conceptions of individuals are influenced by and help create normative values in a society. Whether constructed as a learning or policy problem, greater understanding of teacher beliefs and knowledge is an important aspect of teacher education (Cochran-Smith & Fries, 2005). In addition, since policy changes to methods and purposes of school assessment require the active participation of teachers, our results support the inclusion of prospective teachers' beliefs about the nature and purpose of assessment in research on pre-service teacher preparation programs (Ludlow et al., 2008; Lukin, Bandalos, Eckhout, & Mickelson, 2005; McMunn, McColskey, & Butler, 2003). Teacher education programs certainly would benefit from tools that monitor the effectiveness of programs in developing assessment literate prospective teachers. The current study has suggested that the TCoA is a useful beginning point for developing clearly focused, culturally and societally valid, and appropriate research tools into prospective teacher beliefs about the nature and purpose of assessment.

In addition to standard psychometric evidence offered in this paper, future studies should examine, with non-anonymous data, the effect of teaching experience on conceptions of assessment and the effect of assessment related beliefs on classroom teaching practices. However, in order to demonstrate the relationship of beliefs to practices and outcomes, we must first establish robust measures of teacher beliefs. This study makes a step towards a better self-report inventory and a better understanding of cross-cultural elements related to the psychology of assessment in schools.

## References

- Arbuckle, J. L. (2008). AMOS (Version 17.0.0) [computer program]. Crawfordville, FL: Amos Development Corporation.
- Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). New York, NY: Open University Press.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1–26. <http://dx.doi.org/10.1146/annurev.psych.52.1.1>.
- Betoret, F. D., & Artiga, A. G. (2004). Trainee teachers' conceptions of teaching and learning, classroom layout and exam design. *Educational Studies, 30*(4), 355–372. <http://dx.doi.org/10.1080/0305569042000310309>
- Brown, G. T. L. (2001–2003). *Teachers' conceptions of assessment (TCoA) inventory* (Versions 1–3). Unpublished test. Auckland, NZ: University of Auckland.
- Brown, G. T. L. (2004a). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports, 94*, 1015–1024. <http://dx.doi.org/10.2466/pr0.94.3.1015-1024>
- Brown, G. T. L. (2004b). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301–318. <http://dx.doi.org/10.1080/0969594042000304609>
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99*, 166–170. <http://dx.doi.org/10.2466/pr0.99.1.166-170>
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Brown, G. T. L. (2009). Teachers' self-reported assessment practices and conceptions: Using structural equation modelling to examine measurement and structural models. In T. Teo & M. S. Khine (Eds.), *Structural equation modelling in educational research: Concepts and applications* (pp. 243–266). Rotterdam, NL: SensePublishers.
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation, 6*(12), 68–91.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice, 15*(1), 3–17. <http://dx.doi.org/10.1080/09695940701876003>
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction, 19*(2), 97–111. <http://dx.doi.org/10.1016/j.learninstruc.2008.02.003>
- 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. <http://dx.doi.org/10.1080/09695940903319737>
- Brown, G. T. L., Lake, R., & Matters, G. (2009). Assessment policy and practice effects on New Zealand and Queensland teachers' conceptions of teaching. *Journal of Education for Teaching, 35*(1), 61–75. <http://dx.doi.org/10.1080/02607470802587152>
- 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. <http://dx.doi.org/10.1016/j.tate.2010.08.003>
- Brown, G. T. L., Irving, S. E., & Keegan, P. J. (2008). *An introduction to educational assessment, measurement and evaluation* (1st ed.). Auckland, NZ: Pearson Education.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures - the issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466. <http://dx.doi.org/10.1037%2F0033-2909.105.3.456>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. [http://dx.doi.org/10.1207/S15328007SEM0902\\_5](http://dx.doi.org/10.1207/S15328007SEM0902_5)
- Cochran-Smith, M., & Fries, K. (2005). Researching teacher education in changing times: Paradigms and politics. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Coll, C., & Remesal, A. (2009). Concepciones del profesorado de matemáticas sobre la evaluación del aprendizaje en la educación obligatoria: dimensiones y funciones. [Mathematics teachers' conceptions of learning assessment in compulsory education: dimensions and functions]. *Infancia y Aprendizaje, 32*(3), 391–404.
- Croft, A. C., Strafford, E., & Mapa, L. (2000). *Stocktake/evaluation of existing diagnostic tools in literacy and numeracy in English*. Wellington, NZ: NZCER.
- Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles Policy & Practice, 9*(2), 237–253. <http://dx.doi.org/10.1080/0969594022000001959>
- Crooks, T. J. (2010). Classroom assessment in policy context (New Zealand). In B. McGraw, P. Peterson & E. L. Baker (Eds.), *The international encyclopaedia of education* (3rd ed., pp. 443–448). Oxford, UK: Elsevier.

- Delandshere, G., & Jones, J. H. (1999). Elementary teachers' beliefs about assessment in mathematics. A case of assessment paralysis. *Journal of Curriculum and Supervision*, 14(3), 216–240.
- 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.
- Esparza, E. R. (2007). *Encuesta Abreviada de Concepciones de Evaluación del Aprendizaje*. Unpublished translation of Teachers' Conceptions of Assessment inventory (CoA-III A), University of Sonora, Mexico.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509–529. <http://dx.doi.org/10.1080/00273170701382864>
- Fiske, E. B., & Ladd, H. F. (2000). *When schools compete: A cautionary tale*. Washington, DC: Brookings Institution Press.
- Griffiths, T., Gore, J., & Ladwig, J. (2006, November). *Teachers' fundamental beliefs, commitment to reform and the quality of pedagogy*. Paper presented at Australian Association for Research in Education Annual Conference, Adelaide, Australia.
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365–381. <http://dx.doi.org/10.1080/09695940903319745>
- Hill, M. F. (2000). *Remapping the assessment landscape: primary teachers reconstructing assessment in self-managing schools*. Hamilton, NZ: University of Waikato.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4<sup>th</sup> ed., pp. 17–64). Westport, CT: Praeger.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 18–24.
- Levin, B. (2001). *Reforming education: From origins to outcomes*. London, UK: RoutledgeFalmer.
- Li, W. S., & Hui, S. K. F. (2007). Conceptions of assessment of mainland China college lecturers: A technical paper analyzing the Chinese version of CoA-III. *The Asia-Pacific Education Researcher*, 16(2), 185–198. <http://dx.doi.org/10.3860%2Ftaper.v16i2.262>
- Ludlow, L., Pedulla, J., Enterline, S., Cochran-Smith, M., Loftus, F., Salomon-Fernandez, Y., & Mitescu, E. (2008). From students to teachers: using surveys to build a culture of evidence and inquiry. *European Journal of Teacher Education*, 31(4), 319–337. <http://dx.doi.org/10.1080/02619760802420842>
- Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2005). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practices*, 23(2), 26–32. <http://dx.doi.org/10.1111/j.1745-3992.2004.tb00156.x/pdf>
- MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, issues, and applications* (pp. 16–36). Thousand Oaks, CA: Sage.
- 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. <http://dx.doi.org/10.1207/s15328007sem1103>
- Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.
- McMunn, N., McColskey, W., & Butler, S. (2003). Building teacher capacity in classroom assessment to improve student learning. *International Journal of Educational Policy, Research, & Practice*, 4(4), 25–48.
- Ministerio de Educación Español (2007a). *Orden ECI/2572/2007 Sobre evaluación en Educación Secundaria Obligatoria*. [Law ECI/2572/2007 On assessment in Compulsory Secondary Education] Madrid, Spain: Ministerio de Educación.
- Ministerio de Educación Español (2007b). *Orden ECI/2572/2007 Sobre evaluación en Educación Primaria*. [Law ECI/2572/2007 On assessment in Primary Education] Madrid, Spain: Ministerio de Educación.
- Ministerio de Educación Español (2007c). *Real Decreto 1467/2007 Por el que se establece la estructura del Bachillerato y se fijan sus enseñanzas mínimas*. [Royal Decree 1467/2007 Establishing the structure of High School and its basic curriculum] Madrid, Spain: Ministerio de Educación.
- New Zealand Ministry of Education. (1994). *Assessment: Policy to practice*. Wellington, NZ: Learning Media.
- New Zealand Ministry of Education. (2007). *The New Zealand Curriculum for English-medium teaching and learning in years 1-13*. Wellington, NZ: Learning Media.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), pp. 307–332. <http://dx.doi.org/10.3102/00346543062003307>
- Remesal, A. (2006). *La resolución de problemas en las prácticas de evaluación del área de matemáticas en la educación obligatoria: ideas de profesores y alumnos*. [Problem solving in assessment practices of mathematics education in compulsory school: Ideas of teachers and pupils]. Unpublished doctoral thesis. University of Barcelona, Spain.
- 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, 27–38. <http://dx.doi.org/10.1080/09585170701292133>
- Remesal, A. (2008, August). *Student teachers' conceptions on assessment: First year conceptions*. Paper presented at Northumbria EARLI-SIG Assessment, Potsdam, Germany.
- Remesal, A. (2009). *Cuestionario Abreviado de Concepciones de Evaluación del Aprendizaje*. Unpublished translation of Teachers' Conceptions of Assessment inventory (CoA-III A), Universidad de Barcelona, Spain.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472–482. <http://dx.doi.org/10.1016/J.TATE.2010.09.017>
- van den Berg, B. (2002). Teachers' meanings regarding educational practice. *Review of Educational Research*, 72(4), 577–625. <http://dx.doi.org/10.3102/00346543072004577>

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(4), 4–70. <http://dx.doi.org/10.1177/109442810031002>

Wilkinson, I. A. G., & Townsend, M. A. R. (2000). From Rata to Rimu: Grouping for instruction in best practice New Zealand classrooms. *The Reading Teacher*, 53(6), 460–471.

Wolf, D., Bixby, J., Glenn, H., & Gardner, H. (1991). To use their minds well: Investigating new forms of students' assessment. *Review of Research in Education*, 17, 31–74.

Received February 17, 2010

Revision received February 1, 2011

Accepted March 22, 2011

## APPENDIX A

## TCoA-III A items and factors to prospective teachers Model C: English and Spanish wording

CoA-III A	Statement	Original TCoA	Model C TCoA-TES
3	Assessment is a way to determine how much students have learned from teaching. <i>Evaluar es una forma de identificar cuánto han aprendido los alumnos de lo enseñado.</i>	Improve	Improve
4	Assessment provides feedback to students about their performance. <i>La evaluación informa a los alumnos sobre su aprendizaje y rendimiento.</i>	Improve	Improve
5	Assessment is integrated with teaching practice. <i>La evaluación está integrada en la enseñanza.</i>	Improve	Improve
6	<b>Assessment results are trustworthy.</b> <b><i>Los resultados de la evaluación del aprendizaje son de fiar.</i></b>	<b>Improve</b>	<b>School Accountability</b>
12	Assessment establishes what students have learned. <i>La evaluación determina lo aprendido por los alumnos.</i>	Improve	Improve
13	Assessment feedbacks to students their learning needs. <i>La evaluación informa a los alumnos de sus necesidades de aprendizaje.</i>	Improve	Improve
14	Assessment information modifies ongoing teaching of students. <i>La información que aporta la evaluación del aprendizaje modifica la enseñanza sobre la marcha.</i>	Improve	Improve
15	<b>Assessment results are consistent.</b> <b><i>Los resultados de la evaluación son consistentes, (es decir: un mismo alumno tendrá resultados similares en momentos distintos).</i></b>	<b>Improve</b>	<b>School Accountability</b>
21	Assessment measures students' higher order thinking skills. <i>La evaluación mide las habilidades de pensamiento complejo del alumno.</i>	Improve	Improve
22	Assessment helps students improve their learning. <i>La evaluación ayuda a los alumnos a mejorar su aprendizaje.</i>	Improve	Improve
23	Assessment allows different students to get different instruction. <i>La evaluación permite que algunos alumnos diferentes reciban una enseñanza diferente de los demás.</i>	Improve	Improve
24	<b>Assessment results can be depended on.</b> <b><i>Los resultados de las evaluaciones son fiables, (es decir: miden el rendimiento real del alumno).</i></b>	<b>Improve</b>	<b>School Accountability</b>
7	Assessment forces teachers to teach in a way against their beliefs. <i>La evaluación fuerza a los profesores a enseñar de cierta manera contraria a sus creencias.</i>	Irrelevant	Ignore

CoA-III A	Statement	Original TCoA	Model C TCoA-TES
8	Teachers conduct assessments but make little use of the results. <i>Los profesores evalúan el aprendizaje, pero usan muy poco los resultados de esta evaluación.</i>	Irrelevant	Ignore
9	Assessment results should be treated cautiously because of measurement error. <i>Los resultados de las evaluaciones se deben tratar con precaución por el error de medición.</i>	Irrelevant	Ignore
16	Assessment is unfair to students. <i>La evaluación es injusta para los alumnos.</i>	Irrelevant	Ignore
17	Assessment results are filed & ignored. <i>Los resultados de la evaluación son archivados e ignorados.</i>	Irrelevant	Ignore
18	Teachers should take into account the error and imprecision in all assessment. <i>Los profesores deberían tener en cuenta el error y la imprecisión que tiene toda evaluación.</i>	Irrelevant	Ignore
25	Assessment interferes with teaching. <i>La evaluación interfiere en la enseñanza.</i>	Irrelevant	Bad
26	Assessment has little impact on teaching. <i>La evaluación tiene poca repercusión en la enseñanza.</i>	Irrelevant	Bad
27	Assessment is an imprecise process. <i>La evaluación es un proceso inexacto.</i>	Irrelevant	Bad
1	Assessment provides information on how well schools are doing <i>La evaluación del aprendizaje informa acerca de lo bien que van los centros educativos.</i>	School Accountability	School Accountability
10	Assessment is an accurate indicator of a school's quality. <i>La evaluación es un indicador exacto de la calidad de las instituciones de educación.</i>	School Accountability	School Accountability
19	Assessment is a good way to evaluate a school. <i>La evaluación del aprendizaje es una buena forma de valorar un centro educativo.</i>	School Accountability	School Accountability
2	Assessment places students into categories. <i>La evaluación del aprendizaje distribuye a los alumnos en niveles y categorías.</i>	Student Accountability	Student Accountability
11	Assessment is assigning a grade or level to student work. <i>Evaluar es asignar una calificación al trabajo del alumno.</i>	Student Accountability	Student Accountability
20	<b>Assessment determines if students meet qualifications standards.</b> <i>La evaluación del aprendizaje determina si los alumnos han alcanzado los niveles de competencia o habilidad requeridos.</i>	<b>Student Accountability</b>	<b>Improve</b>

Note. Items in bold belong to different factors than the original model shown in Panel A.