TESTING MODELS OF TEACHERS’ CONCEPTIONS OF ASSESSMENT WITH SAMPLES FROM CYPRUS AND NEW ZEALAND

Michalis P. Michaelides  
*European University of Cyprus*  
Gavin T. L. Brown  
*The Hong Kong Institute of Education*


Correspondence concerning this article should be addressed to Dr Gavin T L Brown, Dept. of Psychological Studies, The Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, NT, HONG KONG SAR or by email to: gtlbrown@ied.edu.hk.

Abstract

Surveys of New Zealand and Queensland primary and secondary teachers, with the *Teachers’ Conceptions of Assessment* (TCoA) inventory, have reported four hierarchical, inter-correlated factors (i.e., assessment for improvement, school accountability, and student accountability, and assessment as irrelevant). A weakness of this research is that it has been conducted only in English in two jurisdictions with very similar policies of low-stakes testing. This paper addresses this by comparing New Zealand results with a Greek-language survey of 249 teachers in Cyprus which also has a relatively low-stakes assessment policy during the compulsory school years. Exploratory factor analysis of the Cyprus data led to an alternative five factor solution with 24 items within two inversely correlated 2nd-order factors (i.e., assessment is positive and negative; $r=-.49$). A multi-group nested invariance confirmatory factor analysis found statistical invariance between the Cyprus and New Zealand data. Mean score differences were small for two improvement oriented conceptions, moderate for assessment is bad, and large for school accountability and ignoring assessment factors. Teachers’ conceptions of assessment linked school accountability to improvement and reflected confidence in teacher professionalism to use assessment for improvement.

Keywords:  
Teacher beliefs; educational assessment; survey research; cross-culture comparison; Cyprus; New Zealand
As long as teachers implement assessment policies in school contexts, the nature and structure of teachers’ beliefs about assessment will matter to how and why assessment is carried out. Brown (2008) has argued that teachers’ conceptions of assessment aggregate into four major beliefs (i.e., assessment is a means 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 to educational practice). Indeed, the improvement, accountability, and irrelevance constructs, 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 tend to identify the improvement of teaching and learning as their dominant purpose for assessment. Further, this factor is only weakly correlated with the notion of grading or evaluating students and more negatively correlated with the conception that assessment is bad, unfair, and inaccurate. In addition, survey research with New Zealand secondary students using the Students’ Conceptions of Assessment inventory has shown that they also believe in improvement as the dominant purpose for assessment (Brown, Irving, Peterson, & Hirschfeld, 2009). Another study showed that the conception that assessment evaluates the student through grading was positively associated with increased learning outcomes (Brown & Hirschfeld, 2008). Hence, it appears that when students agree with being evaluated or graded their grades are higher; however, this practice is not 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). Furthermore, Brown and Harris (2009) have argued that social and cultural priorities express themselves in policies and practices that influence teacher beliefs; hence, it is likely that to the degree countries have different systems, teachers will have different beliefs. Cross-cultural research with the TCoA has suggested that cultural, societal, or linguistic differences impact on teacher beliefs systems. For example, Hong Kong has a strong public examination system and cultural acceptance that examinations lead to enhanced social equity (Cheung, 2008). A survey with a Chinese translation of the TCoA found that teachers strongly associated (r=.91) assessment as improvement with assessment as making students accountable (Brown et al., 2009). Hence, where societies differ in their uses of assessment, one might legitimately expect differences in how teachers conceive of assessment.

1.1. Assessment contexts
1.1.1. Cyprus

The current curriculum of Greek Cypriot Education aims toward a gradual introduction and development of the child in the cognitive domain, value system, psychokinetic and socialization domains, as well as civic issues of the Cypriot society (Cyprus Ministry of Education and Culture, 2002). Assessment is explicitly discussed in the formal curricula as an indispensible part of the teaching-learning process with the purpose of improving practice according to the aims set by the Ministry, with the use of valid and reliable measurements, and avoiding selection or rejection of students through norm-referencing. Standardized and teacher-developed assessments are proposed, in addition to observations, communication, self-assessments, while teachers are encouraged to maintain qualitative notes in the form of diaries (Cyprus Ministry of Education and Culture, 2002).

The Cypriot educational system can be characterized as low-stakes with respect to assessment. In the elementary school, assessment activities are mainly classroom-based.
Teachers are required to maintain records with grades for each student, but they are primarily kept for internal school purposes and monitoring of student achievement and not for formal grade reporting. Testing material is provided by the Ministry of Education for a number of subjects, and teachers may use it along with their own assessment activities. In the gymnasium (grades 7 to 9) testing becomes more formal, students and parents are provided with score reports but assessments are teacher-designed and only the end-of-year exams on core subjects are school-wide. The same is true for lyceums and technical schools (grades 10 to 12). There is no compulsory, large-scale assessment system mandated by the government, although students on an individual basis voluntarily participate in international exams or national competitions in subjects such as mathematics, essay, the sciences etc. A discussion took place in the early 2000s on the introduction of an educational policy with components of standardized testing to monitor school performance, but was never adopted. Accountability at the school level is not a familiar notion for Cypriot teachers, and it may appear appealing in the current discussions about reforming teacher assessment and the evaluation of school performance.

One exception is the final exam system at the end of the 12th grade. These are high-stakes national examinations that certify high-school graduation, and, at the same time, provide access scores for the public universities and tertiary institutions in Cyprus and Greece (Cyprus Ministry of Education and Culture, 2010). The national exams are perceived favourably by students and the public in general, as well as fair and challenging; preparation in the form of private, afternoon tutoring is widely prevalent in the population (Michaelides, 2008). We could hypothesize that teachers will support the idea of assessment as a means for improving teaching and learning, and for holding students accountable; although not implemented in practice, and not generally supported by major stakeholders, assessment practices for school accountability linger as alternative measures in current discussions of educational reform.

1.1.2. 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) relative to the outcomes objectives specified in the national curriculum. The curriculum is child centred, non-prescriptive, holistic, and integrated while, simultaneously, having managerial overtones with specified outcomes and objectives across multiple levels. There is no compulsory, state mandated assessment regime in the primary school sector; hence, all assessment practices are voluntary and low stakes. 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. Hence, we should expect that teachers in New Zealand are strongly committed to the notion of assessment for improved learning and teaching, while not avoiding the use of assessment to evaluate school quality and to grade student learning for certification processes.
1.3 Research Questions

Given the transition that teachers must implement assessment policies and practices in their own schools and classrooms, this study evaluated the measurement model of the Teachers’ Conceptions of Assessment Abridged (TCoA-III) inventory (Brown, 2001-2003) with New Zealand and Greek-Cypriot teachers. The research questions examined were:

1. Does the original TCoA-III model, developed in New Zealand, fit the responses of Greek-Cypriot teachers?
2. What measurement model best describes how Greek-Cypriot teachers respond to the TCoA-III?
3. Is the revised model statistically invariant for samples of practicing teachers from the two countries?
4. What context variables most likely explain the responses of the teachers from the two sites?

2. Method

This study makes use of responses to the Teachers’ Conceptions of Assessment inventory from three independently carried out surveys of practicing teachers in Cyprus and New Zealand.

2.1 Participants

2.1.1. Cyprus

A sample of 249 Greek-speaking Cypriot educators responded to the TCoA inventory. The majority of the participants was female (75%), consistent with the overall composition of the Cyprus teacher body: 81% in elementary and 61% in secondary education (Statistical Service, 2009). Most of the respondents were elementary school teachers (53%), many were secondary school teachers (36%) and the rest worked as preschool teachers. In terms of experience, there was representation from teachers of all levels of experience: novice (none or less than two years of experience) teachers 20%, 2-5 years 17%, 5-10 years 25% and over 10 years of experience 39%.

2.1.2. New Zealand

The New Zealand samples consist of two waves of nationally representative responses. A sample of 525 primary school teachers was collected in 2001 with details reported in Brown (2002, 2004a), while a sample of 404 secondary teachers was collected in 2007 with details reported in Brown (2007). The majority of participants were female (primary 76%; secondary 54%) and the vast majority (83%) were of New Zealand European or related North American, European, South African ethnicities. These proportions are consistent with the 2004 Teacher Census1 which had 80% of respondents as European/Pakeha and 82% of primary teachers and 58% of secondary teachers as female.

2.2 Instruments

The TCoA-III uses 27 statements to construct nine factors, organised into four major 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). Respondents use a six-point, positively-packed, agreement rating scale known to generate discrimination in contexts of social desirability (Brown, 2004b). The responses were coded: 1=Strongly Disagree, 2=Mostly Disagree, 3=Slightly Agree, 4=Moderately Agree, 5=Mostly Agree, 6=Strongly Agree. The TCoA-III 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 Greek and three teachers provided input on the quality and comprehensiveness of the translated items. An English language academic

---

provided a back translation in English which was compared to the original scale and the comparison resulted in minor adjustments in the Greek text. Next, the full scale was administered to two groups of teachers attending graduate level courses in educational research methods. After completing the scale the two groups provided comments about the clarity, and comprehensiveness of the items, and engaged in level-specific (i.e. elementary, and secondary school level) interpretations they gave to the items. Two major issues emerged from this discussion: first, that there are quite distinct conceptions about assessment items between teachers in secondary schools and those working at the elementary or preschool levels; second, it was noted that the term assessment, translated in Greek as \textit{axiologisi} (αξιολόγηση), is used for multiple purposes including the evaluation of student achievement, of teachers, of teaching, and of school outcomes.

In the administration of the TCoA inventory, the teachers were instructed to interpret the term “assessment” according to their subjective conception of the term. In contrast to the original English version, the Greek version employed a balanced 6-point agreement scale. The Greek scale was coded: 1=Completely disagree, 2=Disagree to a large degree, 3=Disagree somewhat, 4=Agree somewhat, 5=Agree to a large degree, 6=Completely agree. Note that insofar as the confirmatory factor analysis is concerned the different meaning of the response values is irrelevant; it is only when scale scores are compared that an adjustment is required.

Six of the respondents, who were English (ESL) teachers, were asked to complete both the Greek and the English version of the inventory. Their mean scores in the four dimensions of assessment conceptions in English were correlated to the corresponding Greek scores and the correlations were strong ($r= .95, .83, .71$), except for School Accountability ($r=.06$). Hence, other than this last scale, given only six cases, there was initial evidence that responses across the two languages were equivalent.

2.3 Analysis

A combination of exploratory and confirmatory factor analyses was used to evaluate the TCoA inventory responses. All cases with more than three 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). 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 1st-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 analysed and cross-validated with the data from a sample drawn from a different population (i.e., Cyprus or New Zealand).

The analytic strategy was first to attempt to fit the original TCoA model previously validated with New Zealand and Queensland teachers (Brown, 2006) to the Cyprus data. Failing that model, a second model that removed all 1st-order factors but retained the four meta-factors was tested. If both \textit{a priori} models failed to fit well to the Cyprus data, exploratory factor analysis with maximum likelihood estimation and oblique rotation was used to detect the most likely model for the Cyprus data which would then be validated with confirmatory factor analysis. Inspection of modification indices and covariance structures was used to determine whether a hierarchical structure could be imposed on the newly generated model. Having found the best-possible model for the Cyprus teacher responses, the
Invariance of the model was tested on the New Zealand sample to establish whether the responses of the two samples could be legitimately compared.

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, \( \text{GAMMA hat} > .90 \), 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 Cyprus 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 the models. Invariance was tested by progressively constraining parameters (i.e., metric invariance of equivalent regression weights from 1st-order factors, then invariance of equivalent regression weights from 2nd-order factors, then invariance of equivalent covariances between 2nd-order factors, then invariance of equivalent residuals for 2nd-order factors, 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, then invariance is imputed (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000).

Invariance of paths, regression weights, and covariances are required to infer the model fits identically in both samples.

Results

Measurement model for Cyprus teachers

Fit statistics based on multi-group confirmatory factor analysis for the four models are in Table 1. The original New Zealand TCoA model (Figure 1, Panel A) was found to be inadmissible due to extremely negative error variance value on the inaccurate first-order factor (\( z = 2.38, p = .02 \)). While the simplified New Zealand model (Figure 1, Panel B) of only four factors predicting 27 items was admissible, the fit failed to meet conventional standards. The five factor EFA result (cumulative variance explained = 41%) removed one item about measurement error because it had no loading greater than .30 on any of the five factors (Figure 1, Panel C). This model had acceptable fit statistics with a very noticeable pattern of inter-factor correlations suggesting a negative and positive dimension in teacher conceptions of assessment. It was also noteworthy that two items had very low regression weights from their respective factors; hence, they were removed in the next analysis. An inter-correlated 2nd-order structure that classified the five factors (with 24 items) according to whether they were positive or negative (Figure 1, Panel D) was tested and a statistically significant improvement in fit over Model C was found (i.e., \( \Delta \chi^2 \) with \( df \) has \( z = 4.67, p < .001 \) [Wilson & Hilferty, 1931]). The fit of Model D is strong enough that the model does not need to be rejected.

Invariance of Cyprus Model D: Comparison to New Zealand Teachers

A three-group (i.e., Cyprus, NZ Primary, NZ Secondary) nested invariance study of Model D was conducted. Because the error variance the Improved Teaching factor was positive for two groups, the negative error variance for the New Zealand Secondary group (\( z = 2.26, p = .02 \)) was resolved by fixing the variance to .005 for all three groups (Chen, Bollen, Paxton, Curran, & Kirby, 2001). Fit across multiple indices was sufficient to support not rejecting the model (Table 1, Model D_{invariance}).

Support for configural invariance was acceptable (i.e., RMSEA=.050 for the unconstrained model). Inspection of the differences in the comparative fit index (CFI) suggested non-chance differences between the groups in their responses to the model. Fixing the 1st-order factor to item regression weights to be equivalent (i.e., metric invariance)
resulted in CFI=.006; fixing the 2\textsuperscript{nd}-order factor to 1\textsuperscript{st}-order regression weights to be equivalent resulted in CFI=.001; fixing the covariance between the two 2\textsuperscript{nd}-order factors to be equivalent produced CFI=.001; fixing the 2\textsuperscript{nd}-order factor residuals to be equivalent resulted in CFI=.000; and fixing the measurement residuals to be equivalent resulted in CFI=.024. Since equivalent measurement residual values are not a requirement of invariance, the evidence of the change in CFI supports the use of Model D to compare the conceptions of these three samples of Cyprus and New Zealand teachers.

Although the rating scales used different verbal anchors, the terms are reasonably close in meaning, except for rating point number 3, where the Greek ‘Disagree Somewhat’ is more clearly negative than the English ‘Slightly Agree’. In order to align meanings before mean score comparisons, all Disagree Somewhat responses in Greek were recoded to 2 ‘Mostly Disagree’ so that the New Zealand and Cyprus scales could be compared. Since there were so few people choosing response 3 (range was 5 to 88 per item), there was little difference in the item level mean scores after recoding (mean effect size was $d=.14$, $SD=.08$). The factor mean scores, standard deviations, and Cronbach’s alpha for the three samples, after recoding the Greek values, are displayed in Table 2.

The Cypriot teachers endorsed all five factors more strongly than either group of New Zealand teachers. The general tendency spills over to the overall Positive factor second which the Cyprus group endorsed considerably more than the two New Zealand groups ($d>.60$). This tendency may suggest a response bias difference between the two countries, wherein Cypriot teachers are simply more inclined to agreeable to statements about assessment.

However, comparing the Cypriot data with the two New Zealand samples, for two of the five 1\textsuperscript{st}-order factors, the mean score differences were trivial (i.e., $d<.20$), one factor had small differences (i.e., $0.21<d<.40$), and two had large differences (i.e., $d>.60$). The least differences were seen around teachers using assessment to improve and assessment is bad; whereas the greatest differences were seen around school accountability and ignoring assessment. The low level of differences around two of the positive conceptions of assessment (i.e., teachers improve teaching and students improve learning) seems to suggest that practicing teachers’ have a relatively universal positive belief around explicit improvement-oriented conceptions of assessment.

In contrast, there were large country differences around endorsing assessment as a means of school accountability and ignoring assessment. In both cases, Cypriot teachers agreed much more with both of these conceptions, notwithstanding their logically contradictory meaning. Consistent with Model D, the Pearson correlations for the Ignore Assessment were negative towards the positive factors and positive towards the Bad factor. Hence, we are unable to tell from these data what aspect of assessment Cypriot teachers believe should be ignored; it is clearly not the three positive purposes identified in this study.

**Interpretation of Model D**

The English and Greek items for Model D, along with regression weights, are displayed in Appendix A. This model suggests that teachers have two major belief systems about assessment (i.e., a positive and a negative one) which are moderately inverse to each other ($r=-.65$, $r^2=.42$). This amount of shared variance suggests that the two systems tend to be bipolar opposites; however, there is a degree of independence in the two factors especially around the Ignore response which has a mean score approaching the positive system. Hence, it would appear that assessment in general is likely to elicit simultaneously both negative and positive attitudes.

The Negative conception consists of two subordinate ideas (i.e., assessment is bad and assessment is ignored). Assessment is bad because it (a) interferes with teaching, (b) is unfair to students, (c) forces teachers to teach against their own beliefs, (d) has little impact on their
teaching, and (e) is an imprecise process. It is worth noting that all of these items were in the same category in the original New Zealand TCoA entitled Irrelevance. The Ignore factor consists of two items (i.e., little use is made of assessment results and assessment results are filed and ignored). This factor appears to capture the teacher response to bad assessments—ignore them.

In contrast, the positive conception consists of three subordinate ideas (i.e., assessment improves teaching, assessment improves student learning, and assessment holds schools accountable). Assessment improves teaching because it (a) helps students improve their learning, (b) is dependable, (c) measures students’ higher order thinking skills, and (d) allows different students to get different instruction. This captures the teacher formative response to assessment—it is carried out to inform the quality of teaching. Assessment improves student learning because it (a) establishes what students have learned, (b) determines how much students have learned from teaching, (c) provides feedback to students about their performance, (d) permits modification of ongoing teaching of students, (e) feedbacks to students their learning needs, (f) is integrated with teaching practice, and (g) determines if students meet qualifications standards. This factor is more student-focused and in line with a student-involved version of formative assessment. It is geared around giving students knowledge about their performance (including relative to external qualifications) for the explicit purpose of improving their learning and the teaching they receive.

Assessment holds schools accountable because it (a) is a good way to evaluate a school, (b) provides information on how well schools are doing, (c) is an accurate indicator of a school’s quality, (d) is trustworthy and consistent, and (e) assigns a grade or level to student work. These items focus on the accuracy of assessment scores and grades as an indicator of provider quality. Without accuracy of course it might not be possible to hold schools accountable through student assessment; but if the scores are accurate then it reflects quality of schooling.

Discussion

This study has investigated the validity of the New Zealand developed questionnaire and analytic model for the Teachers’ Conceptions of Assessment inventory with a sample of Greek-Cypriot teachers. The New Zealand model was found to be inadmissible for the Cypriot teachers and after exploratory factor analysis a hierarchical, inter-correlated model of positive and negative conceptions of assessment was found. Since sample size was less than 300 for the Cyprus sample, it is possible that there was insufficient power to accurately estimate opinions. However, Model D was found to be equally well-fitting and statistically invariant in a multi-group analysis of three groups of teachers (i.e., New Zealand primary and secondary, and Greek-Cypriot). This suggests that the current model is worth interpreting in terms of implications for all three groups of teachers. This model reduces the conceptual complexity of the previous models for the Teachers’ conceptions of assessment inventory (Brown, 2008). Instead of having four inter-correlated factors, the current model proposes that these various purposes can be subsumed under the notion of a positive- and negative-orientation towards assessment. A plausible interpretation is that assessment is considered as indispensible and beneficial in the teaching-learning process; at the same time, teachers are aware of misuses of assessment or policies and practices that touch on their professionalism, i.e. they themselves being evaluated based on their student results. It is worth noting that these two orientations are not strictly opposite to each other; the moderate inter-correlations suggest that teachers have both positive and negative conceptions of assessment simultaneously. Even though they are logically contradictory, there is a sense of plurality in teachers’ conceptions of assessment—logically opposite notions are not fully inverse in the minds of teachers.
In contrast, there is greater support for the validity of the TCoA-IIIA inventory itself. Overall, most items aggregated with their logically consistent factors in this reanalysis. As seen in Appendix A, 12 of the improvement items originated in the improvement conception, the three school accountability items remain within school accountability, and the seven irrelevance items remain in the negative orientation. Hence, 20 items have not changed overall sense in this reanalysis compared to previous studies (Brown, 2004a, 2006, 2007, 2008). The major exceptions appear to be the items to do with the inaccuracy and imprecision of assessment results which do not fit this model at all; perhaps these items refer more to teacher practices than teacher beliefs about assessment. The results suggest that the items in the current questionnaire may form the basis of a universal questionnaire and that the new model is a better specified and more universal model than originally developed. Although, comparisons with other samples of practicing teachers is required for this claim to be more rigorously tested.

The results for four of the factors are entirely consistent with Cyprus’s child-centered educational system that emphasizes the use of low-stakes, teacher-controlled judgments and tests without external, compulsory national testing. Indeed, the positive ratings for Teacher Improvement and Student Learning combined with the negative rating for Bad, suggest that it is this very system that allows Cypriot teachers to endorse positively (and more strongly than New Zealand teachers) the School Accountability factor. In other words, it seems that Cypriot teachers conceive that evaluating schools with assessments is legitimate, since the assessment system and policy are consistent with high respect for and trust in teachers’ professionalism in evaluating, monitoring, and responding to student learning progressions. Brown (2009) has argued that New Zealand primary school teachers associated school accountability with assessment practices that focus on deep cognitive outcomes; whereas, the same teachers associated grading students with assessments that focus on surface cognitive processing. It may be that the Cypriot teachers, likewise, share a conviction that their assessments are a valid basis for evaluating school and teacher performance, since teacher assessments focus on the truly valued learning outcomes of the Cypriot curriculum. Moreover, the lack of centralized policies for school accountability may render this conception attractive given the generally positive stance toward the classroom assessment process.

Differences in mean scores were large around two factors (i.e., school accountability and ignoring assessment) with Greek-Cypriot teachers endorsing these more than both groups of New Zealand teachers. Classroom assessment practices in Cyprus are not standardized across the country, especially for the primary teachers who constitute the majority of the Cypriot sample. Filing and use of results is not mandatory in the elementary grades and external monitoring policies are uncommon; rather this rests with the individual teacher and the guidelines communicated by the school principal, who may or may not instruct the teachers to file assessment results in certain formats. The individual teacher may choose for example to share student results with parents or present them when she is evaluated by the inspector, but that is an entirely subjective teacher practice. The whole approach is further promoted by a general belief that tests and grades may be used, but with care and a critical eye, especially when the students are younger. Alternate, as suggested earlier, the increased score may reflect a response style among Cypriots—further research is needed to explain this result.

The similarity of ratings for Teacher Improvement, Student Learning, and Bad factors suggests that there are some commonalities across societies in terms of teachers’ conceptions of assessment. Both Cyprus and New Zealand share relatively low-stakes assessment policies, especially around school accountability and evaluation. We would argue that the model of beliefs presented here is a positive and constructive set of beliefs that empower
teachers to conduct ‘educational’ assessment; that is, assessment that leads to improved teaching and learning. This is what Popham (2000, p. 1) claims is the only reason to do assessment: “if educational measurement doesn’t lead to better education for students, then we shouldn’t be doing it … the only reason educators ought to assess students is in order to make more defensible educational decisions regarding those students”. Hence, under the current policy in Cyprus there is a strong connection in teachers’ conceptions of school accountability as being related to improvement.

In contrast, since the data reported in this paper were collected, there have been policy changes within New Zealand as to how assessment is being used. National externally-mandated testing to evaluate schools has been introduced (New Zealand Ministry of Education, 2010) and an increasing number of schools are using low-stakes tests to demonstrate school accountability (Brown & Harris, 2009). In the latter context, it has already been found that New Zealand teachers’ conceptions of assessment have become considerably more oriented towards school accountability while being considerably less oriented towards improvement (Brown & Harris, 2009). This suggests the introduction of stronger, external-monitoring policies are not really needed; teachers in both these low-stakes countries are already committed to demonstrating their quality. They may lose a positive orientation towards educational improvement in the face of increased accountability pressures that seem to reduce confidence in teacher professional capabilities.
References


Harris, L. R., & Brown, G. T. L. (2009). [details removed for peer review].


Figure 1. Schematic models of Teachers’ Conceptions of Assessment inventory