



Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Version

This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 <http://www.niso.org/publications/rp/>

Suggested Reference

Brown, G. T., Chaudhry, H., & Dhamija, R. (2014). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research*, 71, 50-64. doi: [10.1016/j.ijer.2015.03.001](https://doi.org/10.1016/j.ijer.2015.03.001)

Copyright

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

© 2015, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

<http://www.elsevier.com/about/company-information/policies/sharing>

<http://www.sherpa.ac.uk/romeo/issn/0883-0355/>

<https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm>

The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools.

Gavin T L Brown

The University of Auckland

Harish Chaudhry

India Institute of Technology-Delhi

Ratna Dhamija

Australian Council for Educational Research (India)

Correspondence should be addressed to Associate Professor Gavin T L Brown, Faculty of Education, The University of Auckland, Private Bag 92019, Auckland, New Zealand 1142 or by email to gt.brown@auckland.ac.nz.

Acknowledgements. This project was supported by the New Zealand India Research Institute (grant #9015/3704567). Data collection and preparation was managed by EduExcellence India and the contributions of Himanshu, Kapil, Chetan, and Naveen are especially appreciated. Data analysis support was provided by Anurag and Shruti, PhD students at IIT Delhi, Department of Management Studies.

Abstract

India has engaged in a policy reform seeking to increase the formative use of assessment in the hope of reducing negative effects of public examinations on students. The 2005 Curriculum Framework has been implemented within the context of significant privatisation of schooling around the country. This study examined the beliefs of teachers about the purpose of assessment because they are the main agents of the policy reform. A large-scale survey of secondary school teachers predominantly in private schools asked them to indicate how much they agreed with multiple purposes concerning either internally-determined school-based assessments ($n=812$) or externally-mandated public examinations ($n=883$) and how they practiced assessment. Structural equation modeling identified a well-fitting model in which there were eight statistically significant paths from Beliefs to Practices and which was strictly equivalent between conditions. While teachers in both conditions endorsed most strongly the improvement purpose, there were statistically significant differences in mean score between conditions for three of the purposes and for one practice. While differences accounted for just 3% of variance in factor means, they were in the hypothesised direction in which internal school-based assessment generated more endorsement of the improvement purpose and diagnostic practice. Greater use of diagnostic practices (an ambition of the Indian Curriculum Framework) depends, in part, on teachers believing in the positive role of internal, school-based assessment and emphasis on educational improvement as the legitimate purpose of assessment is to be encouraged.

Keywords: teacher beliefs; assessment and evaluation; secondary schooling; India; survey research; structural equation modeling

Recommended citation:

Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015, accepted). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research*.

There are various assessment policy reforms globally that seek to address a variety of challenges. The imposition of compulsory student testing to evaluate schools and teachers has been a key characteristic of American educational policy for the last two decades (Ravitch, 2013). In contrast, resistance to the mandated Key Stage testing at ages 7, 9, 11, and 14 in England, led to a strong formative assessment reform advocacy (known as assessment for learning) which has been widely endorsed in many developed Commonwealth countries (Stobart, 2006). Dissatisfaction with examination systems that failed to identify the actual competencies of adolescents and adults (hidden behind rank order scores or examination percentages) had led to the successful implementation outcomes or competence based qualifications (Crooks, 2002). In some societies, there has been a widespread reaction against the reduction of teaching to the development of examination-taking skills, because they are deemed useful for passing examinations that focus on accurate memorisation of academic content. For example, Hong Kong has implemented a new secondary school curriculum and moved the 13th year of schooling into the 1st year of university education in the hope of increasing students' critical thinking and learning ability (Chan, 2010).

As in other countries highly dependent on formal high-stakes public examinations, the 2005 Indian National Curriculum Framework (NCF) (NCERT, 2005) has tried to move the focus of educational assessment from being purely summative public examinations to a more constructivist and formative footing. Specifically, the NCF sought to renew the curriculum by reforming the examination system and reduce psychological pressures upon children and parents, especially in Classes 10 and 12 when high-stakes public examinations were implemented. Rather than classify children as 'pass' or 'fail', the reform sought to use evaluation practices so as to provide greater feedback to learners, extend the range of evaluated capacities to include non-academic curricular outcomes (e.g., thinking skills, leadership, cooperation, sports, arts, etc.), and incorporate teacher judgements throughout the learning process as part of the feedback to parents and children.

This curricular reform of assessment and evaluation practices requires active engagement and understanding by classroom teachers. Hence, it is important to discover if the new policy has had an identifiable impact on teacher beliefs about, attitudes toward, and self-reported practices of assessment.

Teacher Beliefs about Assessment

It is generally agreed that teachers' belief systems about the nature and purposes of a phenomenon (e.g., teaching, learning, or assessment) influence strongly how they teach and what students learn or achieve (Fives & Buehl, 2012). Due to socialization processes, human beliefs seem to be context-dependent (Gao & Watkins, 2002) and appear to be ecologically rational (Rieskamp & Reiser, 2007). This suggests that as government policies shape educational activities, teacher beliefs will reflect the priorities and even tensions present in a society (Brown & Harris, 2009). For example, New Zealand has an assessment policy that focuses predominantly on the formative, diagnostic, and interactive classroom features of assessment (Ministry of Education, 1994) and teachers there are strongly committed to an improvement-oriented purpose for assessment (Brown, 2004, 2011). In contrast, teachers in examination driven societies, such as, Egypt (Gebriel & Brown, 2014) and China (Brown, Hui, Yu, & Kennedy, 2011), are strongly focused on the evaluation of students as the purpose of assessment. Additionally, as policies change, teacher attitudes and beliefs appear to modify in response to a new policy. For instance, Brown and Harris (2009) indicated that as the practice of leader-driven, school-wide data analysis of assessment results was implemented, teacher beliefs moved from being predominantly improvement-oriented to being dominated by the idea that assessment demonstrated school quality.

Two studies have examined explicitly the relationship of teacher beliefs about assessment and teachers' self-reported assessment practices. New Zealand primary school teachers

responses (Brown, 2009) indicated that the more assessment was seen as a way to hold students accountable the more formal, test-like assessment practices were used which were considered to be measures of surface (i.e., recall of facts, details, and information) learning. In contrast, the more assessment was seen as an indicator of school quality, the more teachers reported using measures of deep (i.e., transformational construction of new meanings from material) learning. Additionally, increased use of informal assessment practices (e.g., teacher-student interaction, student self- and peer assessment) was predicted by the belief that assessment was for improvement and that assessment was irrelevant. Together these patterns suggested that teachers believed externally created measures of student accountability only delivered surface learning, while school-based assessment practices led to improvement, especially of deep learning competences.

Similarly, among Hong Kong primary and secondary teachers (Brown, Kennedy, Fok, Chan, & Yu, 2009), in the context of an assessment for learning project, indicated that they used diagnostic and improvement assessment practices (e.g., analysing student strengths and weaknesses, giving formative feedback, and modifying teaching plans) the more they believed that assessment was for improvement. Consistent with the high-stakes consequences for school reputation based on student examination results, teachers reported increased use of practices intended to show the school was doing a good job (e.g., school self-evaluation based on examinations and using exam results as a quality indicator) when they agreed that assessments ought to be for school accountability. Emphasis on student accountability as the purpose of assessment led teachers to specifically prepare students for external examinations (e.g., help students pass exams, teach exam skills, and teach to exam requirements). Finally, the teachers reported sticking to their teaching plans and ignoring exam items in their classes, when they indicated belief that assessment should be ignored.

Together these two studies, in quite contrasting policy jurisdictions (i.e., highly formative vs. highly summative), show that teacher self-reported practices have meaningful alignment with their beliefs as to the purposes of assessment. On the whole, it would seem teachers are very sensitive to the important role that assessment plays in communicating the quality of a school (and by inference themselves) and report using practices that maximise student performance on external measures. At the same time, teachers indicated strong endorsement of the improvement goal of assessment and the use of diagnostic practices and indicated a willingness to modify teaching in response to assessment information. While these studies reflect teacher perceptions and beliefs and lack independent verification of the espoused practices, they also lack explicit comparison of teacher beliefs in response to contrasting conditions of assessment. Within a population of teachers, studies are needed that determine whether externally mandated assessments (e.g., public examinations) and internally administered school-based assessments (e.g., CCE) elicit different beliefs, attitudes, and practices. Hence, the goal of this study is to examine whether teacher self-reported beliefs about the purposes of assessment and their self-reported assessment practices differ according to the type of assessment.

Indian Context

India has a large secondary school system (NUEPA, 2014), having in 2013 almost 240,000 secondary (Class 9-10) and upper secondary (Class 11-12) schools with almost 6,000,000 pupils enrolled. There are just under 2,000,000 teachers in the sector, who are largely highly qualified (i.e., 45% of secondary teachers had postgraduate or higher qualifications, as did 95% of upper secondary teachers). The average pupil to teacher ratio is 31, but there is an average of 50 pupils per classroom, meaning the balance of teachers function in administrative or support roles. Despite its scale, enrolment beyond elementary schooling is not universal; the gross enrolment rate for secondary schooling is 77% and just 52% in upper secondary schooling.

In addition to scale and socio-economic segregation, India's school system is complicated by the Indian response to historic hierarchical and unequal treatment of people according to their caste. Just over three-quarters of all pupils are identified as being members of a protected caste (e.g., scheduled class, scheduled tribe, other backward class) or of Muslim religion (NUEPA, 2014), each of which is given protected rights and status in society. Nonetheless, despite efforts to support minority and under-privileged groups, Indian schooling tends to reproduce the impact of social and economic privilege in the intake of its students (Chudgar & Quin, 2012) and the pedagogical practices of the socially privileged teachers appear to treat the social minority child as having deficits that make him or her uneducable (Nambissan, 2013).

India, being a federal state, places responsibility for education at the state-level. However, given the imperial history of India when education was a national responsibility, national organisations, which are still current, (i.e., Central Boards) arose to administer various secondary school examinations. Thus, schools are associated and draw their curriculum from one of three kinds of school boards. State Boards within each state of India have very similar curricula but are distinguished by offering the state's own language so that about 45-50% of State Boards have Hindi as additional language. Central boards (e.g., Central Board of Secondary Examination, CBSE; Indian Secondary Certificate of Education ISCE; Senior Secondary Certificate, SSC) teach in English only from Class 10 onwards. Central boards have their own distinctive curricula but have similar pedagogical and evaluative processes revolving around high-stakes summative examinations at the end of Class 10 and 12. International boards, whose qualifications are recognised both overseas and within India (e.g., Cambridge International Examination CIE, International Baccalaureate IB), have curricula from an overseas or global authority, and have English as the medium of instruction. Both government and private schools can be members of either State or Central boards, while only private schools are affiliated with International boards.

Private Schooling

India has adopted, from the 1990s, neo-liberal economic shifts which have resulted in the privatisation and marketization of schooling, the withdrawal of government funding to schools, a growing loss of confidence in government schools, and an increase in private schooling (Nambissan & Rao, 2013). At the time of this study, nearly 38% of secondary schools were privately owned and run (NUEPA, 2014), with an additional 17% private with government aid, though the latter could be considered equivalent to government schools in terms of salaries and performance (Kingdon, 2007). While government schools remain free, there has been a well-documented attestation of declining quality and efficiency among government schools, resulting in the development of private and private-public partnership schools (Kingdon, 2007; Nambissan, 2013). A case has been made that private schools provide a superior educational experience at a much lower price (Tooley, Dixon, Gomathi, 2007), with superior achievement rates for children at the end of Class 5 (Pal, 2010), and at the completion of senior secondary school Class 12 pass rates for fully private senior secondary schools is over 90% (Tyagi, 2010).

Nonetheless, private schooling, despite much lower salary rates than government schools (Kingdon, 2007), remains generally out of reach for the majority of Indian families (Härmä, 2011), despite growing parental aspiration to avoid government schools. In response, through a voucher-type scheme, private schools in the primary sector are being compelled and funded directly by government to set aside 25% of all places for students from disadvantaged homes (Government of India, 2009; Kingdon, 2007).

Education Evaluation Reform in India

Prior to the publication of the NCF, a feasibility study in primary school for implementing continuous comprehensive assessment was conducted (Rajput, Tewari, &

Kumar, 2005). The scheme focused on regular and periodic “systematic data collection regarding all aspects of pupils' education-related growth and development for the purposes of decision making” (Rajput et al., 2005, p. 331) and found that teachers, students, and parents considered the scheme to be useful and practicable for assessing children's all-round development. Subsequently, school-based assessment schemes have been introduced especially among private schools; for example, the Central Board of Secondary Examinations (CBSE) introduced Continuous and Comprehensive Evaluation (CCE) in all its member schools in 2009 (described and critiqued in Nawani, 2013).

The Indian assessment reform seeks to incorporate aspects of formative assessment and broadened curricular focus. For example, Mandal (2010) describes how social sciences as a subject contains a broad range of learning objectives and how these can be used to evaluate cognitive, affective, and psychomotor domains of learning. However, in terms of implementation, the CCE policy seems to be better described as cumulative, summative assessment in which frequent and periodic assessments contribute to final examination grades. For example, the CBSE version of CCE has in each half-year, two 10% formative assessments followed by a 30% half-year examination. A similar evaluation was reached by Ashita (2013) based on observations of a government school deploying CCE and interviews with teachers. Hence, notwithstanding the policy goals of the NCF, the reality is that school-based assessment in Indian schools functions less as formative assessment and more like higher education coursework and terminal examinations that summatively contribute to an overall grade.

In addition to examinations of academic content, teachers make professional judgements concerning the broader aspects of the curriculum (e.g., attitude, effort, leadership, personal and social skills, etc.). These teacher judgements are reported alongside the subject examination performances in reports and are incorporated into an overall judgement as to whether a child qualifies for entry into senior secondary or higher education at the end of Class 10 and 12 respectively.

The design of both the NCF and CCE places a great responsibility on teachers to be the key agents of the joint policy reforms; a challenge for most Indian teachers (Aggarwal, & Bhalla, 2012). Thus, understanding their perceptions is critical to an understanding of the effectiveness of the policy. For example, Singhal (2012) reported government school teachers had moderately positive views of the CBSE CCE scheme, despite concerns over difficulties in implementing the scheme, especially due to large class sizes. In response to such concerns, technological solutions for delivering assessments, remediation, and enhanced reporting are being developed and experimented with (e.g., Raman & Nedungadi, 2010). An exploratory observational study of 20 secondary teachers of English language in CBSE schools found that a wide variety of formative assessment practices were being implemented (Chopra & Bhatia, 2014). However, current perception studies are limited in scope and scale and in sophistication of data analysis. This study extends the current work through a large-scale survey of perceptions around assessment intentions and practices and by linking domains through structural equation modeling.

Furthermore, given the need to be responsive to market forces, it is not surprising that CCE is being taken seriously by private schools and is being actively implemented. Thus, research into the differences, if any, of the impact of internal and external assessment practices and policies is more likely to generate meaningful results from private school teachers.

Research Questions & Hypotheses

Based on the Indian context, previous studies examining relationships of beliefs to practices, and principles of ecological rationality in belief systems, three hypotheses were proposed.

- H1. Beliefs about purposes of assessment will predict conceptually aligned uses or practices of assessment;
- H2. Responses and pathways will differ between the external and summative versus internal and formative types of assessment, with greater emphasis on diagnostic and improvement functions under the internal condition; and
- H3. Because CCE contributes to summative evaluation, despite its formative timing before major mid-year and end-of-year examinations and the potential it has to inform changes in teaching, any statistically significant responses between external and internal conditions will have little practical significance.

Methodology

Design

A large-scale survey of teachers' beliefs, perceptions, and espoused practices using a cross-sectional design (teachers of Classes 9-12) was conducted. Sampling of teachers in private schools attempted to ensure adequate inclusion of schools across a number of characteristics. Quasi-experimental assignment of teachers to either internal or external assessment conditions was used to prompt responses.

Experimental Conditions. The research team considered that internal assessments were any methods of collecting data about student learning controlled by the teacher in the classroom context without explicit mention of any consequence. In contrast, external assessments were defined as those administered by formal external examination authorities, suggesting that such assessment would have consequence for both teacher and student. To stimulate responding in each condition, a prompt was included at the beginning of the questionnaire. In the internal condition the questionnaire began with the prompt:

The term "assessment" used in the following statements refers to any act of collecting and interpreting evidence of student learning in terms of knowledge, skills, values and attitudes **USED BY THE TEACHER WITHIN THE CLASSROOM.**

In contrast, the external condition prompt stated:

The term "assessment" used in the following statements refers to any act of collecting and interpreting evidence of student learning in terms of knowledge, skills, values and attitudes **BY EXTERNAL EXAMINATION AUTHORITIES OR BOARDS (e.g., CBSE).**

Note capitals and bold are as displayed on the questionnaire form.

Participants

While efforts were made to ensure diversity of school characteristics, recruitment was through convenience sampling of volunteer schools and volunteer teachers approached by the professional school development agency EduExcellence. This organisation has for the last decade worked extensively in helping improve school leadership and management. Given the problems of public schooling described earlier and the growth of private schooling, it is unsurprising that the vast proportion of cooperating schools have been in the private sector. The sample of schools (Table 1) was dominated by schools in northern India, closest to the New Delhi base of the research team. Furthermore, the sample was predominated by schools affiliated with Central Boards (and most especially the CBSE, which is the largest Central Board in India). All χ^2 tests of the proportion of teachers within each condition by school factors were statistically non-significant indicating the observed difference in distributions were due to chance.

Participants within schools were also volunteers, with random assignment to condition. Consistent with Indian teacher characteristics (Table 1), nearly three-quarters (74%) of all participants were women. Just over half (57%) were highly experienced; almost all were teachers or senior teachers (91%), just over half (58%) taught only in secondary

schools, and teachers of English and science accounted for over half the sample (54%). Likewise, all χ^2 tests of the proportion of teachers in each condition across personal characteristics were statistically non-significant indicating that differences in distributions were due to chance.

Nonetheless, this sample is characterised by highly experienced women teachers working in privately-owned schools operating in an urban area and affiliated with a central board. This means that generalisations from this large sample to Indian government schools cannot be supported and that this study provides insights as to teacher perceptions in Indian private secondary schooling.

Table 1. Teacher and School Demographic Characteristics by Experimental Condition

Teacher Characteristics	condition		School	Condition	
	Internal	External		Internal	External
Sex			Region		
Female	603	649	North	434	425
Male	209	234	South	148	147
			East	131	147
Experience			West	99	164
Less than 5	136	171			
Between 6 to 10	193	226	Board		
More than 10	483	486	State Boards	136	199
Qualification			Central Boards (e.g., CBSE, ICSE)	664	660
Bachelors	116	129	International Boards (e.g., CIE, IB)	12	24
Post-Graduate (Diploma, Certificate)	202	210			
Masters and Doctorate	494	544	School Governance		
Role			Government	40	40
Trainee Teacher	7	23	Private	772	843
Teacher	379	423	School Location		
Senior Teacher	364	382	Urban	593	654
Assistant, Deputy, and/or Principal	62	55	Semi-urban	219	229
Teaching Level			Rural	10	10
Secondary	473	506			
Senior Secondary	322	358			
Both	17	18			
Teaching Subject					
English	216	229			
Mathematics & Accounting	175	182			
Science	231	237			
Social Sciences	163	195			
Other	27	40			

Instruments

Given that English is the second official language of India and that it is a significant medium of instruction in central boards, the survey was administered in English. Questionnaire administrators were available to assist with nuanced meanings by offering translations into Hindi for those teachers who requested help. The complete questionnaire had 67 self-report rating items concerning purposes and practices of assessment.

Teacher Conceptions of Assessment. The TCoA-III inventory is a 27-item, nine-factor self-reported survey that allows teachers to indicate their level of agreement with statements related to four major purposes of assessment. These are: *Improvement*, which refers to the use of assessment to inform changes in teaching practices or student learning processes (e.g., Assessment is a way to determine how much students have learned from teaching); *Student Accountability* which refers to the evaluation, grading, and certification of student performance (e.g., Assessment determines if students meet qualifications standards); *School Accountability* which refers to the use of student assessment or test/examination results to evaluate the quality of teachers and/or schools (e.g., Assessment provides information on how well schools are doing); and *Irrelevance* which is the view that evaluation processes are inadequate, inaccurate, and/or irrelevant to the teachers' ability to improve student learning (e.g., Assessment forces teachers to teach in a way against their beliefs). The TCoA-III model consists of two 2nd-order factors (i.e., Improvement and Irrelevance), which have four and three 1st-order factors respectively (all containing three items), while the two accountability factors each have three items. The four major purposes are inter-correlated. As a multidimensional inventory, there is no single total score; rather there are four sub-scores based on the aggregation of the items contributing to each factor.

The TCoA-III inventory was developed with a large national survey of New Zealand primary teachers (Brown, 2004) and the abridged version (TCoA-IIIa) was validated with large samples of Queensland primary and secondary teachers (Brown, Lake, & Matters, 2011) and a national sample of New Zealand secondary teachers (Brown, 2011). A Chinese translation was used in Hong Kong (Brown, Kennedy, Fok, Chan, & Yu, 2009), a Greek version in Cyprus (Brown & Michaelides, 2011), a Spanish version in Spain (Brown & Remesal, 2011), and an Arabic translation in Egypt (Gebril & Brown, 2014).

In working with the TCoA in Chinese contexts which are dominated by high-stakes public examinations, Brown, Hui, Yu, and Kennedy (2011) found that teachers identified additional purposes for assessment. One of the more salient purposes was that assessment is used to control teachers' pedagogical and curricular practices (e.g., Assessment is used by school leaders to police what teachers do); this was found to be an aspect of the evaluative and accountability purpose of assessment. Given that schooling in India is also strongly dominated by public examinations, it was decided to incorporate this factor into the beliefs questionnaire. Furthermore, two items were added to complement the control items around covering the examination prescription or syllabus and treating the examination as the standard for teaching. This resulted in 33 items in total to do with teacher conceptions of assessment.

Teacher Practices of Assessment (PrAI). The PrAI inventory (Brown et al., 2009) was developed in Hong Kong to identify the degree to which teachers agreed with assessment practices that (a) diagnose student learning needs, (b) prove school quality, (c) prepare students for high-stakes examinations, (d) improve, change, or adapt teaching in response to assessment information, or (e) ignore or treat as irrelevant assessment information. The Hong Kong teachers agreed more with the two improvement practices ('a' and 'd') by small to medium effects over the two accountability purposes (b and c), and by large effects over the ignore assessment (e). Because the inventory had been developed in a public-examination

society, it was deemed appropriate to explore the self-reported assessment practices of Indian teachers.

Nine items were added to the 23 items of the PrAI in order to extend the teaching for examinations factor, including using alternative assessments to tests and examinations as part of normal practice, grading or marking assessments, review of examination performance to resolve discrepancy in performance. An item was added to the irrelevance factor around the relative priority of curriculum completion over preparing for examinations. An item was added to the school accountability factor around use of assessments to stream students. In total, this section had 34 items.

The PrAI has two items that relate to the notions of measurement error or imprecision in assessments which were considered difficult concepts for Indian teachers who seem to work with test or examination raw scores as if they were entirely accurate estimates. The terms ‘test inaccuracy’ and ‘margins of error’ were substituted in hope that these would indicate to teachers that every assessment and examination is an imperfect measure of proficiency or ability.

Response Format. Participants responded by selecting one of six degrees of agreement ratings that best expressed their opinion about each statement. The rating scale used a positively-packed format in which there are two negative categories (i.e., strongly disagree, mostly disagree) and four positive categories (i.e., slightly agree, moderately agree, mostly agree, strongly agree). This response format is beneficial when it is expected participants are positively inclined towards various constructs—a balanced scale cannot provide variance or discrimination when attitudes are very similar, whereas a positively packed scale generates more variance in the positive range (Klockars & Yamagishi, 1988; Lam & Klockars, 1982).

Procedures

Questionnaire responses were collected by four trained data collectors, each completing on average two schools per day. This mechanism of in-person collection ensured that the questionnaires were completed, rather than ignored or lost in the mail. Four teachers per school, randomly assigned to one of the two conditions, were surveyed in time slots convenient to each teacher’s class schedule. This meant as many as four teachers completed the questionnaire at one time, but most commonly two teachers were surveyed in each session. Each teacher, after having an explanation of the project, self-administered the questionnaire making use of the interviewer for clarification as required. While the questionnaire was administered in English, requests for clarification were answered in Hindi by the interviewer only when that was a common language between teacher and interviewer. Approximately 15% of teachers would ask one or two clarification questions, often around the similarity of a new item to a previous one. Each questionnaire administration took approximately 20-25 minutes to complete.

Analysis

Statistical modeling. Because the questionnaire was made up from two pre-existing components having known factorial structures, it was decided to test those solutions first with confirmatory factor analysis. However, because the Indian data failed to adequately fit the Hong Kong and New Zealand models, exploratory factor analysis was carried out.

The procedures described in Courtney (2013) were followed to determine the most likely number of dimensions for each part of the questionnaire. Maximum likelihood estimation with oblique rotation was used in exploratory factor analysis (Costello & Osborne, 2005). Emphasis was put on the number of dimensions identified by the Velicer’s squared MAP and 4th power MAP. Where multiple solutions were recommended, all were tested, with the most theoretically defensible solution being adopted, provided it met conventional standards for factor analysis. A conventional approach was taken to determining the number

of potential factors and their members: factors had to have (1) at least three items which were conceptually aligned, (2) items with regression loadings of $>.30$, and (3) all cross-loadings had to be $<.30$ (Bandalos & Finney, 2010). After identifying the most plausible factor structure for each construct, the exploratory models were evaluated with confirmatory factor analysis.

Confirmatory factor analysis (CFA) tests the fit of a set of pathways within and among factors by utilising the factor patterns, covariance patterns, and residual or error values within a data matrix (Byrne, 2001). In CFA, relationships between variables and latent factors that are not expected are set to zero, while the expected relationships are free to load onto their appropriate factors (Byrne, 2001). Large samples, usually >500 , are required to provide stable parameter estimates (Chou & Bentler, 1995), which was exceeded by both conditions. All modeling was done in AMOS (IBM, 2011). Although the inventory elicits responses using a six-point, ordinal agreement scale, maximum likelihood estimation with Pearson product moments was used since scales of this length can be treated as continuous variables (Finney & DiStefano, 2006). After finding a robust model for the Beliefs and Practices components separately, a structural equation model (SEM) was developed to identify the prediction from beliefs about assessment to practices of assessment.

There are many measures to assess the fit of a model to the data. The quality of fit for a model to the underlying data matrix is best tested with measures that are not affected by sample size or model complexity; unfortunately, the χ^2 statistic falsely punishes models with large sample sizes; the comparative fit index (CFI) falsely punishes complex models; and the root mean square error of approximation (RMSEA) falsely rewards complex models (Fan & Sivo, 2007). In line with current practice (Cheung & Rensvold, 2002; Fan & Sivo, 2007; Marsh, Hau, & Wen, 2004; Vandenberg & Lance, 2000), acceptable fit for a model was imputed when the χ^2 per df was statistically nonsignificant ($p > .05$), gamma hat $> .90$, and RMSEA and standardized root mean residuals (SRMR) were both $< .08$. Models that met these criteria were retained.

Invariance testing. A feature of CFA and SEM is that they permit examination of whether the parameter values of a model between two or more groups vary by more than chance. If the parameter values are statistically equivalent or invariant, then it can be argued that any differences in factor scores are attributable to differences in the populations from which the samples were drawn, rather than due to deficiencies of the measurement model or inventory (Vandenberg & Lance, 2000). The conventional sequence of equivalence testing, depending on model characteristics, establishes:

- (1) all paths are identical (configural equivalence),
- (2) all regressions from factors to items are equivalent (metric equivalence),
- (3) all intercepts of item loadings on factors are equivalent (scalar equivalence),
- (4) all regressions from factors to other factors are equivalent,
- (5) all covariances between inter-correlated factors are equivalent,
- (6) all structural residuals are equivalent, and
- (7) all measurement residuals are equivalent (strict equivalence).

Scalar equivalence is normally needed to before mean score comparisons between groups can be made (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000): There is general consensus that statistical equivalence between groups does not require structural or item residuals to be equivalent (Wu, Li, & Zumbo, 2002). Further, when scalar invariance is demonstrated, we can conclude that the groups are members of the same population (Cheung & Rensvold, 2002; Wu, Li, & Zumbo, 2007). More importantly, any differences in factor mean scores between the groups cannot be attributed to differential impact of a self-report inventory on participant responses.

The invariance of both measurement and structural models was tested using a nested, multi-group approach (Cheung & Rensvold, 2002). Testing stopped when a parameter was shown not to be equivalent. The configural equivalence of the pathways was accepted if the RMSEA for a multigroup model was $\leq .05$. Differences in the comparative fit index (ΔCFI) should be $\leq .01$ to accept that the additional constraint fits the data (Cheung & Rensvold, 2002; Wu, Li, & Zumbo, 2007). Once at least scalar equivalence was established it was possible to conduct multiple analysis of variance of factor mean scores to establish the extent to which teachers in each condition gave different levels of endorsement.

Results

Teacher Conceptions of Assessment

The nine-factor TCoA model for 27 items was rejected because of negative error variances in three 1st-order factors. Removing all 1st order factors led to acceptable fit ($k=27$; $\chi^2=2335.98$; $df=318$; $\chi^2/df=7.35$; CFI=.81; RMSEA=.06 (90%CI .06-.06); SRMR=.06; γ hat =.92) but this excluded the Chinese control factor. Thus, dimensionality analysis (Courtney, 2013) of all 33 items was used to identify a plausible model for the Indian context. Solutions containing between two (Velicer MAP²) and seven (Spearman CD) factors were systematically evaluated with MLE and oblimin rotation. The two and three factor solutions were rejected because they did not differentiate within the conceptually different aspects of accountability or improvement. The five to seven factor solutions were rejected for failing to ensure that all factors had at least three items with loadings $> .30$ or ensure that there was no conceptual overlap between factors.

The four identified purpose factors were: (a) improvement, which refers to using assessment to identify student learning strengths and needs and provide feedback on those needs; (b) irrelevance, in which assessments are conducted by little use is made of them in determining what a teacher does next in the classroom; (c) control, which refers to the use of assessment to control the teachers' lessons and teaching, usually by focusing on external examination requirements; and (d) school quality, which uses assessment results as a proxy for or indicator of school quality (Appendix A).

A four factor measurement model with the four inter-correlated factors was tested in CFA, producing acceptable fit ($k=26$; $\chi^2=2254.88$; $df=293$; $\chi^2/df=7.70$; CFI=.82; RMSEA=.06 (90%CI .06-.07); SRMR=.06; γ hat =.93). Invariance testing between the two conditions produced statistically equivalent structural covariances ($\Delta CFI \leq .01$) with good fit ($k=52$; $\chi^2=2894.22$; $df=644$; $\chi^2/df=4.49$; CFI=.80; RMSEA=.05 (90%CI .04-.05); SRMR=.07; γ hat =.94).

Teacher Practices of Assessment

Dimensionality analysis (Courtney, 2013) of all 33 items identified between 4 (Velicer MAP²) and 7 (Spearman CD) factors each of which was systematically inspected in EFA with MLE and oblimin rotation. Solutions of five to seven factors were rejected because one or more factors failed to meet the convention of at least 3 items loading $> .30$. The four practice factors identified were: (a) diagnostic, in which teachers use assessment to analyse student needs and teaching effect; (b) school evaluation, in which the school, rather than teacher, uses assessment results to determine its public reputation and to sort students into classes; (c) teaching for exams, in which all types of assessments, including alternatives to tests and exams, are used to prepare students for performance on public examinations; and (d) ignore exams, in which the teacher prioritises their pre-existing teaching plans (rather than examinations) either because they don't have time or consider exams to be inaccurate or bad (Appendix B.)

The four factor inter-correlated solution (29 items) consisting of Diagnostic; School Evaluation; Teaching for Exams; and Ignore Exams was evaluated in CFA. Low factor loadings and/or high modification indices identified four items that were not well specified

and after their deletion acceptable to good fit was obtained ($k=25$; $\chi^2=1887.16$; $df=269$; $\chi^2/df=7.02$; CFI=.88; RMSEA=.06 (90%CI .06-.06); SRMR=.05; gamma hat =.93). Invariance testing between low and high-stakes conditions demonstrated that strict equivalence of equivalent measurement residuals was obtained between the two conditions.

Relationship of Conceptions to Practices

Inspection of the factor inter-correlation matrix (Table 2) showed weak to moderate correlations within Purposes and Practices (average absolute inter-correlation for Purposes was $r=.28$; for Practices $r=.34$). The average between construct correlation was likewise weak (average absolute $r=.25$); although the value for the conceptually similar constructs was stronger (average absolute $r=.38$). This indicated that, within each construct, the factors were sufficiently distinguished from each other and that there was a generally stronger association between purposes and practices that were aligned with each other.

Table 2. Inter-correlations between and within Domains

Domain and Factor	Purposes				Practices			
	1	2	3	4	5	6	7	8
Purpose Beliefs								
1. Improve	—							
2. Irrelevant	-.313**	—						
3. Control	.232**	.114**	—					
4. School Quality	.566**	-.104**	.330**	—				
Practice Beliefs								
5. Diagnostic	<i>.572**</i>	<i>-.208**</i>	<i>.222**</i>	<i>.424**</i>	—			
6. School Evaluation	<i>.298**</i>	<i>-.084**</i>	<i>.333**</i>	<i>.389**</i>	.434**	—		
7. Teaching for Exams	<i>.433**</i>	<i>-.146**</i>	<i>.355**</i>	<i>.341**</i>	.509**	.430**	—	
8. Ignore Exams	<i>.099**</i>	<i>.192**</i>	<i>.196**</i>	<i>.240**</i>	.192**	.318**	.131**	—

Note. Values in bold show within construct inter-correlations; values in italics show inter-construct correlations; values in red show conceptually aligned factors; $N=1695$; $**=p<.01$.

To fully evaluate the systematic relationship of assessment purposes to assessment practices a structural equation model was developed. The four predictor purpose factors were inter-correlated, as were the four dependent practices factors. To allow practices to be correlated as dependent variables, residuals for each factor were introduced and inter-correlated. This is exactly equivalent to having the actual factors inter-correlated. Regression paths were first drawn from each Purpose factor to its conceptually equivalent Practice factor. Then paths from each Purpose Factor to all other Practices factors were introduced and trimmed if they were not statistically significant. The resulting model had eight statistically significant paths from Purposes to Practices. Nested invariance testing showed that ΔCFI was $\leq .01$ for all parameters up to equivalent measurement residuals making the model strictly equivalent between the two conditions of questionnaire administration. Fit indices ranged between acceptable and good ($k=104$; $\chi^2=8834.87$; $df=2582$; $\chi^2/df=3.29$ ($p=.07$); CFI=.78; RMSEA=.04 (90%CI .04-.04); SRMR=.06; gamma hat =.93).

Of a possible 16 paths from Purpose factors to Practices factors, ten were statistically significant (Table 3). The strongest paths were from two of the conceptually aligned purpose-practice combinations (i.e., Improvement to Diagnostic and School Quality to School Evaluation). The diagnostic practice was predicted strongly by Improvement and weakly by School Quality purposes; Teaching for Exams was moderately predicted by both Improvement and Control purposes, with a weak inverse contribution from the Irrelevant purpose; School Evaluation practices was moderately predicted by School Quality and

weakly by Control purposes; while Ignoring Exams was weakly predicted by Irrelevant, School Quality, and Control purposes. The proportion of variance explained in each Practice factor by these relationships was large (i.e., $f^2 > .35$, Cohen, 1992), except for Ignoring Exams which only had a moderate effect.

Table 3. Purposes to Practices Regression Weights

Purposes	Practices			
	Diagnostic	Ignore Exams	School Evaluation	Teaching for Exams
Improve	.60			.38
Irrelevant		.14		-.11
School Quality	.12	.25	.40	
Control		.14	.22	.32
SMC (R^2)	.47	.13	.31	.38
Effect f^2	.89	.15	.45	.61

Note. Conceptually aligned factors shown in bold.

Mean Score Differences

Given that the sufficient invariance was demonstrated, mean scores for the eight factors could be compared between the two conditions. Mean scores were calculated by averaging the response for all items predicted by each latent trait (Table 4). This permits interpretation of factor means on the same response scale as used by the teachers in evaluating each item. Multiple analysis of variance for the eight scales was tested for experimental condition as the sole fixed factor.

Table 4. Factor mean scores and comparison statistics by experimental condition

Statistic	TCOA					Practices		
	Improve	Ignore	Control	School Quality	Diagnose	School Quality	Teach for Exams	Ignore Exams
External								
<i>M</i>	4.74	2.82	4.21	4.05	4.65	4.50	4.85	3.31
<i>SD</i>	.744	.87	.89	.89	.90	1.02	.62	1.07
Internal								
<i>M</i>	4.88	2.76	3.97	4.18	4.94	4.43	4.80	3.26
<i>SD</i>	.69	.86	.95	.93	.77	.95	.59	1.08
Comparison Statistics (External vs. Internal)								
<i>F</i>	17.275	2.309	29.164	9.119	48.969	1.986	2.383	0.795
<i>p</i>	<.001	0.129	<.001	0.003	<.001	0.159	0.123	0.373
$R^2_{(ADJ)}$	0.01	0.001	0.02	0.01	0.03	0.001	0.001	<.001
<i>d</i>	-.20	.07	.26	-.15	-.34	.07	.08	.04

Note: $N(\text{external})=883$; $N(\text{internal})=812$; negative d = Internal is higher than External

Teachers in both conditions endorsed most strongly the Improvement purpose, while Teaching for Exams was the most strongly agreed practice in the external condition and Diagnostic practice was most agreed in the internal condition. There were statistically significant differences in mean score between conditions for three of the purposes (except Ignore) and only for one practice (i.e., Diagnostic). The internal condition also elicited greater agreement that assessment was for school quality and less agreement that assessment controlled teaching. While these differences accounted at best for 3% of variance, reflecting trivial to small effect sizes, they were in the hypothesised direction in which internal school-

based assessment generated more endorsement of the Improvement purpose and Diagnostic practice.

Discussion

The implications of the experimental treatment are discussed first, followed by a comparison of the factor structures of the two inventories relative to their sources. We conclude with some considered speculations as to the implications of the study for future research and educational policy. As well, we speculate as to the social and cultural origins of Indian teacher conceptions of assessment.

Formative Evaluation in Indian Schools

Despite stimulation of participants to consider either internal or external types of assessment, the statistical equivalence of responding and the small scale of mean score differences, leads us to conclude that teachers in private secondary schools in India have fundamentally similar perceptions as to the purposes and practices of assessment. Nonetheless, there was a statistically significant, albeit small, trend in which the internal assessment condition was associated with higher endorsement of the improvement purpose and greater espoused use of diagnostic practices of assessment. Furthermore, the strongest association was between endorsement of improvement as the predictor of diagnostic practices. It would appear, then, that attention to the improvement purpose of assessment is likely to lead to greater use of diagnostic practices (i.e., changing teaching because of what we learned about students from the assessment). To the extent that this reflects the ambitions of the Indian NCF, then this positive impact of internal, school-based assessment and emphasis on educational improvement as the legitimate purpose of assessment is to be encouraged.

However, it seems teachers, regardless of internal or external conditions still see assessment predominantly around improving student learning by teaching for exams. It is possible that the condition prompts were not powerful to ensure distinction between internal and external in teacher responding. Nevertheless, it is clear that CCE is not actually being implemented in a purely formative fashion; each assessment, despite its formative timing, is used predominantly as a cumulative, summative evaluation. Hence, it is highly likely that no distinction between internal and external assessment conditions is a logical consequence of the use of CCE as a contributor to external board-related certification decisions.

Nonetheless, given the ambitions of the NCF and the small trend towards using internal assessments diagnostically, it would seem possible to take advantage of this willingness to be formative with assessments. Even if the teachers in this study have provided responses that are simply repeating official policy, the majority has still indicated endorsement of improvement purposes and diagnostic practices. This provides an important policy and practice lever for policy makers. Teachers want to make a difference and believe assessment can and should contribute to that. At the same time this study identified that there is a complicated relationship between having time to finish the curriculum and teaching plans while using assessment diagnostically.

This means new resources are needed by teachers. Formative CCE testing has to provide diagnostic information to the classroom teacher about who needs to be taught what next; rather than simply total score or rank order information (Brown & Hattie, 2012). However, if this process is not automated through appropriate software, it would be extremely unrealistic to expect teachers to carry out such work manually especially given class sizes and workloads in India (Hattie, Brown, & Keegan, 2003). Furthermore, if the NCF policy is to be properly implemented, not all school-based assessments need to be graded and contribute to summative consequences; teachers need the psychological safety to discover that their best efforts have failed and support from school leadership to discover what new teaching materials or techniques might lead to better results (Brown, 2012). A wider range of

assessment tools are needed independent of cumulative grading so that not every assessment event counts for student grades (Hattie & Brown, 2008). Research has shown that when teachers have access to formative assessments that are supportive of teacher workloads, teachers can and do use even summative assessments formatively (Archer & Brown, 2013; Carless, 2011).

Evaluating the Indian Questionnaire Model

Within the domain of assessment purposes, the Improvement and Irrelevance purposes in this study were populated entirely by items from the original New Zealand TCoA-III inventory related to the same two factors. The School Quality purpose, in contrast, contained the three original School Quality items, two items from Improvement, and one from Student Accountability. The Control purpose builds on two items from the Chinese School and Teacher Control factor and one from the Examination factor. The two new items exhibit similar characteristics that assessments, and especially examinations, are used to control the curriculum and teaching that teachers implement. Given the high consequences attached to school-based assessments and the ease of using examinations as a control mechanism, this factor seems logically coherent with the Chinese teacher perceptions of the purposes of assessment.

Within the Practices domain there are some interesting similarities and differences in the current result relative to the Hong Kong Assessment Practices Inventory. The Indian Diagnostic practices consists of six items from the Diagnose and Improvement factors consistent with the claim made by Brown et al. (2009) that these two factors related to an improvement orientation toward assessment. Likewise, the School Evaluation practice retained all three of the items identified by Hong Kong teachers as focused on this practice. In addition to the five new items developed by the authors, the Indian Teaching for Exams factor was made up of four Examination Preparation and four Improvement items from the Hong Kong version. The Indian Ignore Exams factor consisted of two items from the Irrelevance factor in Hong Kong and were supplemented by a Hong Kong Examination Preparation item and an original item. While there are some differences to the Hong Kong results, the overall impression is that Indian teachers, like Hong Kong teachers, report using assessment to prepare students for examination and that this is considered an important facet of improvement.

It is a possibility that the lack of differences between conditions is a consequence of the failure of the prompts to satisfactorily focus teacher attention on the different type of assessment. Perhaps, teachers simply responded to the notion of assessment rather than gave any weight to the bolding in the prime. Future investigations could substitute the words examination or CCE in place of assessment to test this threat.

Origins of Indian Teacher Assessment Conceptions

The current survey portrays Indian teacher perceptions of the purposes of assessment and its implementation in ways that seem consistent with the previous Hong Kong and Chinese studies in which the quality of a school is indicated by student examination performance and by improvement in that performance. Simply put, it seems that teachers in quite diverse contexts believe a good school's effect is seen on better examination performance. The similarity of perceptions between Hong Kong and Indian teachers may be attributable to the similarity of their working in high-stakes, public examination evaluation systems. Both environments are characterised by competitive and limited rewards (e.g., entry to higher education) based on the merits of performance on formal examination. Both jurisdictions have inherited a British (and perhaps more so an English) model of education that has relied on public examinations for sorting or tracking students into and within schools and for determining access to higher education.

However, historical colonisation is not a sole explanation for the similarity in responses. As outlined previously (Brown et al. 2009; 2011) Confucian values contribute to the strong commitment among Chinese teachers to using examinations validly to improve student learning and personal character. However, this philosophic framework clearly does not apply in this context. Traditional (i.e., pre-British colonisation) Indian approaches to teaching and assessment were based on the interpersonal, apprenticeship experience of living with a guru and graduation was based on the judgment of the individual guru as to the fitness of a disciple to teach independently. Hence, the large-scale, industrial approach to schooling with formal common examinations was not a norm, suggesting that historic patterns in schooling do not necessarily explain the commitment to examinations seen in contemporary India. Nonetheless, commitment to literacy and scholarly knowledge (e.g., Vedic scriptures and science in contemporary India) has been long held as important values.

In contrast, Indian society has long been defined by caste and clan characteristics (easily determined by inspection of family name) which operated so that individual life chances were proscribed and determined by social origins of one's family. One's future career, occupation, income, social status was determined at birth in a society with little social mobility. The introduction of public examinations, where performance rather than social origins were determinant, produced meritocratic social mobility. Hence, for teachers, especially those from non-upper caste backgrounds, endorsement of examinations as a positive force for improvement and quality seems logical. In seeking to modernise itself, India has moved to break down social origins as the basis of selection, promotion, or privilege and bring about a more democratic and meritocratic society; educational examinations are a powerful mechanism by which talent and ability can be identified and rewarded.

Hence, confidence in examinations to bring fair and equitable results seems to be an appropriate response provided the examination system actually brings about social mobility and change. However, given the conservative and social reproductive role of schooling and examinations (i.e., children of the privileged groups generally do better), it is possible that the meritocratic ambition of the examination system is in fact an illusion. The loss of confidence in the public system and faith in private sector solutions may actually undermine social change. However, it is the goal of EduExcellence and other similar organisations to bring about greater life chances for all children, not just those of the better castes and clans. Nonetheless, greater implementation of CCE seems to be having a positive impact on teacher thinking towards better quality information and this may be just a small-stepping stone towards a better Indian schooling.

Conclusion

This study extends smaller previous studies into Indian teacher perceptions of internal school-based assessment by specifically focusing on teachers working in the cutting edge of Indian schooling; that is, teachers in private schools affiliated predominantly with the CBSE. This study shows that these private-school teachers have positive attitudes towards the NCF curriculum goals of broadening attention to multiple learning domains and using internal school-based assessment for the intended formative goals. Nonetheless, the study reveals that perceptions of assessment are equivalent between internal school-based and external examination conditions. Certainly, changes have to be made to the operation of internal assessment if teachers are to see differences between internal assessments that are diagnostic and formative and external assessments that are evaluative and summative.

However, it needs to be kept in mind that this study reports perceptions of teachers about their beliefs and practices. There is no independent evidence in the study as to what is actually happening. It could be that teachers are using the summative 10% within term tests formatively by analysing the content of the student performance in light of the test content

and making adjustments to teaching plans. Such a strategy would be consistent with Carless (2011) recommendations in Hong Kong to use summative testing formatively. However, direct observation of teacher practice, inspection of the trace documents related to their analysis and feedback from CCE events, and perceptions and experience information from students themselves would all help to establish the implementation fidelity of these espoused beliefs.

This study shows that, at least among private school teachers, there is a realistic basis for believing that the policy of internal school based assessment has had a desirable effect on teachers' perceptions. It is this possibility that needs to be extended by policy makers and assessment developers. Teachers want educational assessment, not just evaluation; and it is up to schools, boards, and funders to support such goals.

References

- Aggarwal, S., & Bhalla, V. (2012). Continuous and comprehensive evaluation: Redefining the role of teachers. *Educational Research, 10*, 82-90.
- Ashita, R. (2013). Beyond testing and grading: Using assessment to improve teaching-learning. *Research Journal of Educational Sciences, 1*(1), 2-7.
- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York: Routledge.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice, 11*(3), 301-318.
- Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*, 45-70.
- Brown, G. T. L. (2012). School-based assessment. In H. Chaudhry (Ed.), *Transformational leadership: Some ideas from leading practitioners—Proceedings of the IIT Education Conference 2011 (Excellence in School Education III)* (pp. 64-71). New Delhi, India: Scholastic India.
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources engender heightened conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation, 6*(12), 68-91.
- Brown, G. T. L., & Hattie, J. A. (2012). The benefits of regular standardized assessment in childhood education: Guiding improved instruction and learning. In S. Suggate & E. Reese (Eds.) *Contemporary educational debates in childhood education and development* (pp. 287-292). London: Routledge.
- Brown, G. T. L., & Michaelides, M. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education, 26*(3), 319-337. doi: 10.1007/s10212-010-0052-3.
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology, 15*(1), 75-89. doi: 10.5209/rev_SJOP.2012.v15.n1.37286
- Brown, G. T. L., Hui, S. K. F., Yu, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*(5-6), 307-320. doi: 10.1016/j.ijer.2011.10.003
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice, 16*(3), 347-363. doi: 10.1080/09695940903319737
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*(1), 210-220. doi: 10.1016/j.tate.2010.08.003
- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: LEA.
- Chan, W. (2010). A Review of Educational Reform – New Senior Secondary (NSS) Education in Hong Kong. *International Education Studies, 3*(4), 26-35.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

- Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 37-55). Thousand Oaks, CA: Sage.
- Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review, 31*, 376-390. doi: 10.1016/j.econedurev.2011.12.003
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), Available online: <http://tinyurl.com/557zra>.
- Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment Research & Evaluation, 18*(8), Available online: <http://pareonline.net/getvn.asp?v=18&n=18>.
- Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy & Practice, 9*(2), 237-253.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509–529.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. D. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age Publishing.
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham & T. Urdan (Eds.), *APA Educational Psychology Handbook: Individual Differences and Cultural and Contextual Factors* (Vol. 2, pp. 471-499). Washington, DC: APA.
- Gao, L., & Watkins, D. A. (2002). Conceptions of teaching held by school science teachers in P.R. China: Identification and cross-cultural comparisons. *International Journal of Science Education, 24*(1), 61-79.
- Gebril, A., & Brown, G. T. L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian Teachers' Conceptions of Assessment. *Assessment in Education: Principles, Policy and Practice, 21*(1), 16-33. doi: 10.1080/0969594X.2013.831030
- Government of India. (2009). *The Right of Children to Free and Compulsory Education Act, 2009*. New Delhi: The Gazette of India Extraordinary Retrieved 21 September, 2014 from http://righttoeducation.in/sites/default/files/Right%20of%20Children%20to%20Free%20and%20Compulsory%20Education%20Act%202009%20%28English%29%20%28Copiable_Searchable%29.pdf#overlay-context=.
- Härmä, J. (2011). Low cost private schooling in India: Is it pro poor and equitable? *International Journal of Educational Development, 31*, 350–356. doi: 10.1016/j.ijedudev.2011.01.003
- Hattie, J. A. C., & Brown, G. T. L. (2008). Technology for school-based assessment and assessment for learning: Development principles from New Zealand. *Journal of Educational Technology Systems, 36*(2), 189-201.
- Hattie, J. A. C., Brown, G. T. L., & Keegan, P. J. (2003). A national teacher-managed, curriculum-based assessment system: Assessment Tools for Teaching & Learning asTTle). *International Journal of Learning, 10*, 771-778.
- IBM. (2011). Amos [computer program] (Version 20, Build 817). Meadville, PA: Amos Development Corporation.

- Kingdon, G. G. (2007). The progress of school education in India. *Oxford Review of Economic Policy*, 23(2), 168-195.
- Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement*, 25(2), 85-96.
- Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement*, 19(4), 317-322.
- Mandal, P. K. (2010). Towards positing a paradigm for continuous and comprehensive evaluation in Social Sciences. *Journal of Indian Education*, 36(3), 43-52.
- 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.
- Nambissan, G. B. (2013). Opening up the black box? Sociologists and the study of schooling in India. In G. B. Nambissan & S. S. Rao (Eds.), *Sociology of education in India: Changing contours and emerging concerns* (pp. 83-102). New Delhi, IN: Oxford University Press.
- Nambissan, G. B., & Rao, S. S. (2013). Introduction: Sociology of education in India—Trajectory, location, and concerns. In G. B. Nambissan & S. S. Rao (Eds.), *Sociology of education in India: Changing contours and emerging concerns* (pp. 1-23). New Delhi, IN: Oxford University Press.
- Nawani, D. (2013). Continuously and comprehensively evaluating children. *Economic & Political Weekly*, 48(2), 33-40.
- NCERT. (2005). *National Curriculum Framework 2005*. New Delhi, India: National Council of Educational Research and Training.
- Pal, S. (2010). Public infrastructure, location of private schools and primary school attainment in an emerging economy. *Economics of Education Review*, 29, 783-794. doi: 10.1016/j.econedurev.2010.02.002
- Rajput, S., Tewari, A. D., & Kumar, S. (2005). Feasibility study of continuous comprehensive assessment of primary students. *Studies in Educational Evaluation*, 31, 328-346. doi: 10.1016/j.stueduc.2005.11.002
- Raman, R., & Nedungadi, P. (2010, September). *Adaptive learning methodologies to support reforms in continuous formative evaluation*. Paper presented at the 2010 International Conference on Educational and Information Technology (ICEIT 2010), Chongqing, PRC.
- Ravitch, D. (2013). *Reign of Error: The hoax of the privatization movement and the danger to America's public schools*. New York: A. E. Knopf.
- Rieskamp, J., & Reimer, T. (2007). Ecological rationality. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 273-275). Thousand Oaks, CA: Sage.
- Singhal, P. (2012). Continuous and comprehensive evaluation: A study of teachers' perceptions. *Delhi Business Review*, 13(1), 81-99.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 133-146). London: Sage.
- Tooley, J., Dixon, P., & Gomathi, S. V. (2007). Private schools and the millennium development goal of universal primary education: A census and comparative survey in Hyderabad, India. *Oxford Review of Education*, 33(5), 539-560.
- Tyagi, R. S. (2010). School-based instructional supervision and the effective professional development of teachers. *Compare: A Journal of Comparative and International Education*, 40(1), 111-125. doi: 10.1080/03057920902909485

- 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.
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), Available online: <http://pareonline.net/getvn.asp?v=12&n=13>.

Appendix A. Teacher Conceptions of Assessment (India) Factors and Statements

Factor & Statement	Source
Improvement	
q13 Assessment feeds back to students their learning needs	TCoA-I
q3 Assessment is a way to determine how much students have learned from teaching	TCoA-I
q4 Assessment provides feedback to students about their performance	TCoA-I
q12 Assessment establishes what students have learned	TCoA-I
q22 Assessment helps students improve their learning	TCoA-I
q14 Assessment information modifies ongoing teaching of students	TCoA-I
q5 Assessment is integrated with teaching practice	TCoA-I
q6 Assessment results are trustworthy	TCoA-I
Irrelevance	
q17 Assessment results are filed & ignored	TCoA-Ir
q8 Teachers conduct assessments but make little use of the results	TCoA-Ir
q16 Assessment is unfair to students	TCoA-Ir
q7 Assessment forces teachers to teach in a way against their beliefs	TCoA-Ir
q27 Assessment is an imprecise process	TCoA-Ir
q25 Assessment interferes with teaching	TCoA-Ir
q26 Assessment has little impact on teaching	TCoA-Ir
Control	
q30 Assessment ensures teachers teach to the defined examination standard	
q31 Assessment controls the content of teachers' classes	TCoA(C)-E
q32 Assessment ensures teachers cover the whole curriculum	
q29 Assessment results contribute to teachers' appraisals	TCoA(C)-C
q28 Assessment is used by school leaders to police what teachers do	TCoA(C)-C
School Quality	
q19 Assessment is a good way to evaluate a school	TCoA-S
q10 Assessment is an accurate indicator of a school's quality	TCoA-S
q20 Assessment determines if students meet qualifications standards	TCoA-St
q1 Assessment provides information on how well schools are doing	TCoA-S
q21 Assessment measures students' higher order thinking skills	TCoA-I
q15 Assessment results are consistent	TCoA-I

Note. TCoA = Teacher Conceptions of Assessment-III Abridged (Brown, 2001-2003); TCoA(C)=item taken from Teacher Conceptions of Assessment (Chinese) (Brown, Hui, Yu, & Kennedy, 2011); TCoA-I=Improvement; TCoA-Ir=Irrelevance; TCoA-S=School Accountability; TCoA-St=Student Accountability; TCoA(C)-C = Teacher & School Control; TCoA(C)-E = Examinations; items not marked are original to the Indian research team.

Appendix B. Teacher Practices of Assessment Inventory (India)

Factors and Statements	Source
Diagnostic	
q53 I use assessment to establish what students have learnt.	PrAI-D
q52 I use assessment to determine how much students have learnt from teaching	PrAI-D
q54 I use assessment to identify student strengths and weaknesses.	PrAI-D
q55 I use assessment to identify students' learning needs.	PrAI-I
q51 I use assessment results to predict future student performance.	PrAI-I
q36 I always use assessment to help students to learn.	PrAI-I
School Evaluation	
q58 My school uses assessment results to determine if students meet standards.	PrAI-S
q59 My school uses assessment results to show how well it is doing.	PrAI-S
q57 My school regards assessment result as an important indicator of school's quality.	PrAI-S
q60 My school uses assessment results to stream students.	
q56 My school evaluates its performance mainly by public examination results.	PrAI-S
Teaching for Exams	
q34 I always set tests and examinations with reference to public examinations	PrAI-E
q50 I use alternative assessments to assess different student abilities.	
q49 I use alternative assessment together with tests and examinations in assessment process.	
q39 I assign a grade or mark to student work as significant part of assessment.	TCoA(C) -E
q40 I design different instruction for different students based on assessment results.	PrAI-I
q48 I teach my students examination skills from time to time	PrAI-E
q33 I always provide feedback to students about their performance.	PrAI-I
q37 I ask questions in class mainly to check students' understanding.	PrAI-I
q46 I teach according to public examinations' requirements.	PrAI-E
q63 Our school puts most effort in preparing students for public examinations	
q38 I ask students to do simulated high-stakes examination exercises.	PrAI-E
q45 I take into account error and imprecision when using assessment results.	
q62 On discussing any inconsistency in students' assessment results, I will review their exam papers.	
q44 I re-teach because students get poor assessment results.	PrAI-I
Ignore Exams	
q66 The priority of my work is to complete the curriculum	
q65 The priority of my work is to help students to pass their examinations.	PrAI-E
q35 I always stick to teaching plan irrespective of poor assessment results.	PrAI-Ir
q42 I do not have enough time to explain assessment items after the test.	PrAI-Ir

Note. PrAI = items taken from Practices of Assessment Inventory (Brown, Kennedy, Fok, Chan, & Yu, 2009); TCoA(C)=item taken from Teacher Conceptions of Assessment

(Chinese) (Brown, Hui, Yu, & Kennedy, 2011); PrAI-D = Diagnose; PrAI-I=Improvement; PrAI-S=School Accountability; PrAI-E=Examination Preparation; PrAI-Ir=Irrelevance; TCoA(C)-E=Examination; items not marked are original to the Indian research team.