High-stakes examination preparation that controls teaching: Chinese prospective

teachers' conceptions of excellent teaching and assessment

Abstract

How prospective teachers conceive of teaching excellence and assessment purposes probably influences how teaching and assessment practices are implemented in the future. This study evaluated, in four normal universities in the People's Republic of China, 765 prospective teachers' responses to two self-report instruments regarding the nature of excellent teaching and the purpose of assessment. Both questionnaires had previously been developed in Mandarin Chinese and validated with large samples of practicing teachers in China. The original models could not be recovered and with exploratory and confirmatory factor analysis alternative models were found. Excellent teaching was conceived as four inter-correlated factors (i.e., Professional, Model, Examination, and Life-long), as also were the purposes of assessment (i.e., Diagnose & Formative, Irrelevant, Control, and Life Character). Structural equation modeling showed that the strongest relationship between teaching excellence and assessment began with the examination factor which positively predicted assessment as Irrelevant and for Life Character development and negatively the Diagnose and Formative purpose. Results are consistent with the high-stakes examination system of China and the status of prospective teachers who have only recently stopped being students.

Keywords

Prospective teachers; Chinese education; beliefs and attitudes; teaching; assessment

Introduction

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). Since government policies shape educational activities, it is expected teacher beliefs will reflect the priorities and even tensions present in a society. Furthermore, since teachers and students have different roles in education (i.e., instructor vs. learner) with different insights into the intent of educational policies and systems, it is expected there will be differences between practicing teachers with teaching experience and prospective teachers about excellent teaching and assessment purposes and contrasts those with the beliefs of practicing teachers in China.

The education system of China places great emphasis on regular high-stakes public examinations, such as the Entrance Examination for Senior High School (zhong kao) and the Entrance Examination for Higher Education (gao kao). Both Confucian tradition and contemporary policy place considerable value on high examination scores for both students and teachers (Li, 2009; Min, 1997; Wang, 1996). Nonetheless, the government of China has recently called for assessment reforms that move evaluation systems away from transmission and memorization of 'bookish' knowledge for purely ranking or selection purposes towards more formative, authentic, and humanistic approaches to assessment (OECD, 2011) including the introduction of integrated quality assessment which emphasizes judging students' personal character (Liu & Qi, 2005; Ministry of Education, 2005). However, as Wang (1996) made clear, concern for all-round development of good character and good person attributes has been a priority in modern China's curriculum since the mid-1950s' focus on the 'Three Goods' (i.e., ideology and morality, study, and physical health) and the 'Five Loves' (i.e., motherland, people, labour, science, and socialism) of the 1980s. Furthermore, official curricular attempts to reduce the domineering impact of examinations in China date back to the 1990s (Han & Yang, 2001). Thus, China has a policy framework and system that press towards two different ends; that is, evaluation and improvement. Hence, it is expected that these conflicting pressures might make it difficult for prospective teachers to think like experienced practicing teachers.

This paper tested the validity of pre-existing statistical models of practicing teacher responses to those of a large sample of prospective teachers and examined the relationship of excellent teaching beliefs to conceptions of assessment. It was expected that prospective teachers would have different belief structures to those of practicing teachers. Further, it was expected that the examination factor within excellent teaching would be a strong predictor of beliefs about the purposes of assessment. On the presumption that contemporary teacher education in China foregrounds the importance of formative assessment practices in line with priorities of the New Basic Curriculum, it was expected that teacher education students would conceive of assessment predominantly around formative assessment, much like practicing teachers.

Conceptions of excellent teaching

Teaching excellence in China seems to have strong Confucian features; for example, excellent teachers should be completely devoted (hui ren bu juan) and act as a role model with caring for the students (Louie, 1984). More recently, educational policy in China has aimed at improving teaching quality and promoting student achievement (Ding, 2010; Feng, 2006). The 2001 New Curriculum reforms sought to improve teacher quality so as to maximize students' all-round development. The reforms aimed to reduce the traditional

emphasis on rote memorization, drilling, monotonous classroom environments, and the heavy burden of homework assignments. Instead, the reforms emphasized a more student-centered teaching style to stimulate students to be active masters of their own learning, engage in inquiry and discussion, and become imbued with a love of learning, self-confidence, self-discipline, and cooperation. These reforms attempt to transform the current Chinese education through improved teaching quality (Hughes & Yuan, 2005).

Field studies with Chinese teachers have found this strong student-centered and examination priority. Cortazzi and Jin (1996) found that excellent teachers loved their job and children, had professional morality and responsibility, good knowledge, and acted as a model for students. Watkins and Zhang (2006) concluded that an excellent Chinese teacher had a deep level of general knowledge, deep knowledge of a particular subject, focused on knowledge delivery, helped students to do well in examinations, cared about student's personal problems, had close relationships with students, was a good moral guide, and promoted positive attitudes to society. Similar priorities have been found in both survey and qualitative studies with middle school teachers in China (Chen, 2007; Chen, Brown, Hattie, & Millward, 2012).

Conceptions of assessment

Current educational policy reform movements attempt to reduce the negative consequences of highly selective educational assessments by placing a greater emphasis on using assessment to inform teaching and learning improvements (Berry, 2011). While improvement is the natural goal of education, it should be noted that accountability mechanisms tend to elicit endorsement of the views of the evaluators (Lerner & Tetlock, 1999). Hence, the greater the accountability pressure within an assessment system (e.g., high-stakes consequences for schools or students), the more likely school-based assessments are to be conceived as fulfilling accountability or preparing for examination purposes. While educational policies may formally advocate improvement-oriented assessment practices, the presence of accountability assessments may subvert such policy intentions. Indeed, a recent analysis of Hong Kong's assessment policy suggests that the assessment for learning direction is a soft policy option (i.e., it requires voluntary compliance by teachers and schools), relative to the more powerful hard policy option of selective qualifications examinations for students which are subsequently used to judge teachers and schools (Kennedy, Chan, & Fok, 2011).

China has a long tradition of high-stakes examinations to select students for limited opportunities in higher levels of education or in higher-ranked educational institutions. There is a long history of social support behind the use of public examinations as a selection tool (China Civilisation Centre, 2007; Han & Yang, 2001). Even in today's China, formal assessment mechanisms are still used extensively to select students into elite schools at all levels of schooling (Gao & Watkins, 2001; Watkins & Biggs, 2001). Teachers within Confucian heritage societies appear to see frequent summative assessment and practice for formal examinations as a means of motivating effort and as a means of guiding instruction (Kennedy, Chan, Fok, & Yu, 2008). An additional important distinction in the function of assessment in Chinese contexts is that a good person is one who scores well because examination results reflect the quality and worth of the individual (China Civilisation Centre, 2007; Li, 2009). Thus, in Chinese thinking, assessment improves the child, their performance, and their virtue.

A recent survey of practicing middle school teacher in China (Chen, et al., 2012) showed that preparing students for examinations was only weakly related to their beliefs about the nature of excellent teaching. However, preparing students for examinations was strongly related to teachers' self-reported classroom practices of controlling students and

5

testing them. Similarly, Deng and Carless (2010) reported that the presence of high-stakes external examinations impeded the use of a task-based, communicative language pedagogy in early primary school teaching. Another survey study of practicing teachers in South China and Hong Kong (Brown, Hui, Yu, & Kennedy, 2011) showed that Chinese teachers strongly associated using assessment for accountability and improvement purposes, although the former were considered somewhat irrelevant, while the latter was considered relevant. In summarizing a series of studies carried out by Chinese graduate students, Gao and Kennedy (2011) reported six major beliefs present among practicing teachers in China; that is, 1) assessment controls schools, teachers, and students; 2) assessment checks that students comply with examination requirements; 3) assessment improves teaching performance; 4) assessment is inaccurate and ignored. All these studies suggest that China's teachers are very much aware of the controlling importance of their students doing well on examinations, but tend to consider such an emphasis as distinct from excellent teaching which focuses on caring for the student as a whole person as well as a learner.

Contrasting beliefs of prospective and practicing teachers

Practicing and prospective teachers are at quite different stages of the educational process. Practicing teachers, by necessity, take professional responsibility for what goes on in the classroom and control those activities to achieve intended goals. They are responsible for the full range of curricular outcomes including caring for children who are less successful in the examination system. In contrast, prospective teachers have not yet made the transition to being an instructor and are normally recent successful graduates of the school qualifications and certification system. While prospective teachers have made a commitment to becoming teachers, they lack the experience of a practicing teacher. It is one of the fundamental goals of teacher education to help prospective teachers develop knowledge, skills, and beliefs about their activities (e.g., teaching and assessment) appropriate for effective functioning in the classroom. However, it is highly likely that, despite curricular intentions to move education away from a strict examination society, prospective teachers will have beliefs about educational activities shaped by their own experiences as students in the examination system (Pajares, 1992). Thus, it seems reasonable to expect that prospective teachers will have different priorities in their beliefs than practicing teachers.

Unsurprisingly, over 90% of the 53 Chinese prospective teachers surveyed in one Shanghai teacher education institution considered traditional testing to be necessary, while 42% endorsed the importance of teaching moral development, with personal learning experiences being the dominant source of these beliefs (He, Levin, & Li, 2011). Elsewhere in the world, the beliefs of prospective teachers have been found to differ from those of practicing teachers. This has been especially noted around the purposes and practices of assessment in New Zealand (Hill, Gilmore, Smith, & Cowie, 2012; Smith, Cowie, Gilmore, & Hill, 2012) and in Spain (Brown & Remesal, 2012).

Method

Prospective teacher education in China

In the Chinese teacher education system, prospective teachers are trained at 'normal' colleges or universities (Lin & Xun, 2001). The 'normal' university in China is a multi-disciplinary university which grows out of a dedicated teacher training college. China, unlike many other countries, has no difficulty in recruiting teacher education students for all levels of the education system and student teachers in China are typically academic high-achievers. This may be because teaching is a relatively well-paid job, employment is reasonably certain upon graduation, and has a relatively high social status in China (OECD, 2011). Further, there are few financial barriers to becoming a teacher since tuition is either free or reduced. Since even primary school teaching is 'specialist' in nature (i.e., teachers teach specific subjects in which they have academic qualifications), teacher education constantly balances the mutual demands of ensuring prospective teachers develop both pedagogical skill and subject or discipline knowledge (Xie, 2001). Nonetheless, it is expected in China that much of what teachers need to know and be able to do will be taught through school-based professional learning (OECD, 2011), in accordance with a virtuoso-apprenticeship model of teacher development in which expert, older teachers guide new, younger teachers into the craft and values of teaching (Tsui & Wong, 2009).

Instruments

The instruments used in this study were the Teachers' Conceptions of Excellent Teaching (TCET) (Chen et al., 2012) and the Teachers' Conceptions of Assessment in Chinese Contexts (TCoA-C) (Brown, Hui, Yu, & Kennedy, 2011). Both questionnaires were developed in Mandarin Chinese and validated with large samples of practicing teachers in the People's Republic of China.

The TCET comprises 58 statements which reduce to five factors spread across two broad dimensions about the nature of excellent teaching in Chinese schools. The pedagogical-interactive dimension comprised four factors including developing lifelong learners, student focused, being responsible for engaging students in learning, and being a professional learner, while, the examination dimension was a single factor that was weakly correlated with the pedagogical-interactive dimension.

The TCoA-C comprises 64 statements from which 31 items were extracted in seven factors, which aggregated into three dimensions. The first dimension (i.e., improvement) contains factors related to improving student character, improving student learning, and the

accuracy of assessment. The second dimension was accountability (positively correlated with improvement, r=.80) containing factors related to examinations, control of schools, teachers, and students through assessment, and the inaccuracy of assessment. The third dimension was irrelevance (positively correlated with accountability, r=.22; negatively correlated with improvement, r=-.28) which contains items related to negative effects of assessment and tendency to ignore it.

Responses to both questionnaires indicated how strongly the statements were endorsed using a six-point, positively-packed agreement rating scale. Positively packed rating scales are known to generate discrimination in contexts of social desirability (Brown, 2004; Lam & Klockars, 1982), which is especially expected within Chinese psychology (Bond & Hwang, 1986). There were two negative options (i.e., strongly disagree=1, mostly disagree=2) and four positive options (i.e., slightly agree=3, moderately agree=4, mostly agree=5, and strongly agree=6). Since one of the TCET factors focused on preparing students to do well on examinations, a meaningful and statistically significant relationship to the purposes of assessment was expected from that factor.

Participants

On a convenience basis, the authors contacted the research office at four universities in China and, after briefing each head as to the project, obtained permission to recruit volunteer participants within each university. Once each head agreed to participate, volunteers from among the lecturers in the department were recruited. Lecturers were asked to distribute Student Participant Information Sheets and the questionnaire to the students in their classes. Students were asked to return completed questionnaires within four weeks directly to the research team using pre-addressed, stamped envelopes or to a drop-box on campus.

From the four normal universities approached, 765 questionnaires were returned. Just

over half (n=448, 58.6%) were from women. About two-fifths of the sample were in their 1st year (n=328, 42.9%), almost a third (n=248, 32.4%) were in their 2nd year, and a quarter (n=189, 24.7%) were in their 3rd year of university study. About two-fifths (n=309, 40.4%) were training for primary school teaching, around one-third (n=256, 33.5%) were training for middle school, and one-quarter (n=200, 26.1%) were training for secondary school. About half (n=398, 52.0%) were studying sciences (i.e., chemistry, computers, technology, mathematics, and physics) while the balance (n=367, 48.0%) were studying social sciences and humanities (i.e., arts, Chinese, English, music, geography, history, and politics). Note that no 4th year students participated as these students are normally in school-based teaching practicum and not on campus.

Analysis

All cases with more than 10% missing responses were removed and any remaining missing values were imputed using the expectation maximisation procedure (Little & Rubin, 2002), resulting in no missing data. Confirmatory factor analysis was used to determine whether the responses of the participants fit the pre-existing factor models for the TCET and TCoA-C. Upon discovering poor-fit for each model, exploratory factor analysis was used to develop an alternative teacher education (TE) model for each inventory (TE-TCET and TE-TCoA-C respectively), and confirmatory approaches were used to establish the fit of the new trimmed model. Maximum likelihood estimation with oblique rotation was used in exploratory factor analysis (Costello & Osborne, 2005). A conventional approach was taken to determining the number of potential factors and their members: factors had to have (1) eigen-values>1.00, (2) at least three items which were conceptually aligned, (3) items with regression loadings of >.30, and (4) all cross-loadings had to be <.30 (Bandalos & Finney, 2010). Modification indices were also used to identify and remove items with strong cross-factor loadings.

Structural equation modeling was then used to determine the relationship of the TCET to the factors of the TCoA-C. Predictor paths were tested from each TCET factor to the TCoA-C factors and statistically non-significant paths were removed. Note details of steps taken to develop revised models are omitted for reasons of space.

In line with current practice (Cheung & Rensvold, 2002; Fan & Sivo, 2007; Marsh, Hau, & Wen, 2004), the criteria for acceptable model fit were: (1) statistically non-significant χ^2 per *df*, (2) gamma hat >.90, and (3) both root mean square errors of approximation (RMSEA) and standardized root mean residuals (SRMR) <.08. Models that met these criteria were not rejected. All analyses were carried out in AMOS (IBM, 2011) using Pearson product moment correlations. Note all items with negative factor loadings were reverse scored before mean scores were determined as the average of all items loading on the factor. Participant characteristics (i.e., sex, year of experience in teacher education, and educational level of teacher training) were examined as possible sources of variance in mean scores.

Results

The original models were either inadmissible or very poor fitting, indicating that there were significant differences in responses to the inventories between prospective and practicing teachers. Hence, we report a completely new analysis for prospective teachers. Table 1 provides sample items and descriptive statistics for the excellent teaching and assessment purposes factors, with factor inter-correlations in Table 2.

Excellent teaching

The revised TE-TCET model for teacher education students consisted of four inter-correlated factors (i.e., Life-long, Exams, Model, and Professional) based on 20 items (χ^2 =641.08; *df*=164; χ^2/df =3.91, *p*=.05; CFI=.90; gamma hat=.94; RMSEA=.062; SRMR=.062) (see the

left hand factors in Figure 1). The Life-long factor (made up of 9 items) focused on the responsibility teachers have to help all students learn (including skills, content, character, and attitudes), even through after-hours visits and phone-calls home, and on sensitive, effective in-class teaching practices. The Exam factor in excellent teaching consisted of three items (i.e., matches content taught to what tests measure; spends a lot of class time drilling students with exam-type items; and assesses student learning according to what has been really taught), suggesting that teacher-based assessment must be aligned to examinations and prepare students for success. The Model factor (likewise three items) refers to fulfilling moral obligations by not exploiting their position to make money, being a holistic model for students, and helping students remember more. The Professional factor (five items) refers to teachers' habits of mind, including openness, challenging teaching, appropriate goals, being friendly, and staying current with knowledge and theory.

The Professional mean score had large effect size differences to all other factors (i.e., Exams, d=1.14, Life-long, d=1.67, and Model, d=1.44) (Table 1). Multivariate analysis of variance (MANOVA) with main and all two-way effects for sex, grade within program (i.e., Years 1, 2, and 3), and level for which they were being trained (i.e., primary, middle, or secondary school) found that only one main characteristic (Grade: $F_{8,1498} = 3.32$, p < .01) and one interaction (Grade*Level: $F_{16,2285.82} = 2.186$, p < .01) had statistically significant mean score differences. Univariate analysis showed that the Grade effect applied only to Model ($F_2 = 5.87$, p < .01) and Professional ($F_2 = 3.79$, p = .02), while the Grade*Level interaction only applied to Professional ($F_4 = 3.49$, p < .01). Tukey post-hoc HSD test showed that students in Year 2 were different for Model to those in Years 1 and 3, but there was no distinguishable difference for Professional. While the difference in means was statistically significant, it is more appropriate to conclude that these differences are either practically small or else do not provide a meaningful insight into patterns of difference. Hence, this study, perhaps because

of uneven numbers in each cell, is unable to address student demographic differences in mean scores for the ET factors.

<insert Table 1 about here>

The Examination factor was negatively and moderately correlated with the Life-long factor, and more-or-less independent of the other two factors (Table 2). The Professional factor moderately and inversely related to the Model factor, independent of the Exam factor, and positively but weakly related to the Life-long factor. The Life-long factor was inversely but moderately correlated with the Model factor. These correlations suggest Exams were inverse to Life-long, while Professional was inverse to Model.

<insert Table 2 about here>

Assessment purposes

The revised TE-TCoA-C model had four factors (i.e., Diagnose & Formative, Irrelevant, Control, and Life Character) based on 19 items. Note that three of the TCoA-C factors were dependent on the Diagnose & Formative factor (χ^2 =421.51; *df*=150; χ^2/df =2.81, *p*=.09; CFI=.95; gamma hat=.96; RMSEA=.049; SRMR=.035) (see the right hand factors in Figure 1). The Diagnose factor refers to using assessment to understand student learning so as to guide appropriate teaching responses. This factor negatively predicted responses to the three other factors: Control (i.e., using assessment to control teachers' work), Irrelevant (i.e., assessment is unfair and ignored), and Life Character (i.e., using assessment to help students develop character and life-long learning skills).

The students endorsed Control the most with large effect size differences to all three factors (i.e., Diagnose, d=.85; Life Character, d=1.84; Irrelevant, d=1.98) (Table 1). Note that the mean for factors Irrelevant and Life Character were both less than 3.00 (i.e., slight agreement), suggesting that, on average, these purposes were rejected, rather than endorsed. Multivariate analysis of variance (MANOVA) with main and all two-way effects for sex, grade within program (i.e., Years 1, 2, and 3), and level for which they were being trained (i.e., primary, middle, or secondary school) found that only one main characteristic (Grade: $F_{8,1498} = 2.53, p < .01$) and one interaction (Grade*Level: $F_{16,2285.82} = 1.79, p < .01$) had statistically significant mean differences in agreement with the four conceptions of assessment purposes. Univariate analysis showed that the Grade effect applied only to Diagnose ($F_2 = 3.70$, p = .03) and Irrelevant ($F_2 = 3.58$, p = .03), while the Grade*Level interaction had no statistically significant effects. Tukey post-hoc HSD test showed that only students in Year 2 and Year 3 were different (the latter being lower) for both factors; though effect sizes for this comparison were small (d=.22) to moderate (d=.40). While the difference in means was statistically significant, it is more appropriate to conclude that these differences according to demographic characteristics are sufficiently small they can be disregarded. Hence, this study, perhaps because of uneven numbers in each cell, is unable to address sub-group differences in mean scores for the ET factors.

The factor inter-correlations were all moderate to strong (.50 < |r| < .81), with Diagnose being inversely correlated with the three other factors (Table 2). The three other factors are positively inter-correlated. This suggests strongly that the formative, diagnostic role of assessment was seen as being strongly opposed to the controlling power of assessment which in turn was irrelevant and focused on life-long moral development.

Structural model

The inter-battery correlations (Table 2) had the strongest relations between the Exams factor and the four assessment purposes factor (.50 < |r| < .63). Note that the relationship to Diagnose was negative, while the values were positive otherwise. The structural model (Figure 1), having acceptable to good fit (χ^2 =1999.95; *df*=691; χ^2/df =2.89, *p*=.09; CFI=.89; gamma hat=.92; RMSEA=.050; SRMR=.080), showed that the TE-TCET Exam factor predicted three of the TE-TCoA-C factors (i.e., positive paths to Irrelevant and Life Character and negative path to Diagnose & Formative).

<Insert Figure 1 about here>

The Exam factor within Excellence strongly but inversely predicted the Diagnostic purpose of assessment (β =-.76, R^2 =.57) indicating clearly that concentrating the quality of teaching on preparing students for examinations meant not identifying and responding to student strengths and weaknesses. This interpretation is supported by the weak positive path to Irrelevance (β =.20). The moderate path to Life Character (β =.40) generated a small amount of additional variance explained (additional R^2 =.05). Thus, student teachers' conceptions of excellent teaching for examinations predicted their conceptions of assessment as not formative.

Discussion

As expected, the responses of Chinese prospective teachers to two inventories related to the qualities of excellent teaching and assessment purposes were different to previous studies with practicing teachers in China. Instead of examination preparation being weakly related to excellent teaching, for prospective teachers, it was strongly correlated with caring for students. Instead of a tri-partite understanding of assessment around irrelevance,

accountability, and improvement, the prospective teachers conceived as assessment primarily around its diagnostic and formative functions which were not for control or irrelevant. These differences are consistent with previous studies which have shown that prospective teachers have different belief structures than practicing teachers.

However, the results seem to be consistent with the strong examination culture of China. Excellent teaching was seen predominantly in terms of professional responsibilities of expertise in subject disciplines and care for students, while assessment was seen predominantly as a mechanism that controls, evaluates, and has consequences for teachers. This appears to be an entirely rational understanding of the impact the Chinese public examination system has on students and teachers. Interestingly, the students gave considerably less regard to curricular priorities that require teachers to be honourable models for students and the development of complex relationships with students in which teachers and students are joint learners and in which teachers help students become effective life-long learners. These prospective teachers seem not to have imbibed from the deeply Confucian notions of excellent teaching which were exhibited by practicing middle school teachers in Liaoning province (Chen, 2007).

While factor scores for excellent teaching cannot be directly compared since different items make up the constructs, it is interesting to consider the rank order differences in levels of agreement between this group and the practicing middle school teachers surveyed with the same inventory (Chen et al., 2012). Life-long learning was the most endorsed by the practicing teachers, it was the third most endorsed by prospective teachers. In contrast, prospective teachers gave Exams the second level of endorsement, while it was the least endorsed factor among practicing teachers.

Approaching scores for assessment purposes in a similar way, the prospective teachers' priorities were not identical to those of practicing teachers (Brown et al., 2011). For example,

Irrelevance was the least endorsed factor for both groups, while factors related to helping learning and developing student character were given middle levels of agreement by prospective teachers and were in the top half of priority for practicing teachers. In contrast, Control was the most endorsed by the prospective teachers, while among practicing teachers it was the sixth of seven factors. Thus, it is concluded that prospective teacher views of assessment are that it is very much a constraint on educational activity; whereas among practicing teachers assessment appears to serve more socially endorsed values of developing learning and character.

Hence, it would appear, insofar as beliefs about excellence in teaching and purposes of assessment are concerned, that practicing teachers previously surveyed have beliefs much more in accordance with the New Basic curriculum than these of prospective teachers. Nonetheless, it may be that the practicing teacher responses reflect espoused beliefs that do not reflect enacted priorities and that, potentially, the prospective teachers more accurately reflect priorities as actually being carried out in China's schools.

A major distinctive result of this study is the structure of the Exam factor within excellence teaching beliefs and its role in predicting responses to assessment purposes. Examination was not independent of excellence for prospective teachers (compare Chen et al., 2012); it was inversely related to Life-long and Model factors, but positively related to Professional, suggesting that a highly competent teacher still helps students do well on evaluative assessments; a highly rational response given the importance of examination scores in China. Further, Exams were a negative predictor of the formative uses of assessment. As far as these prospective teachers were concerned, they considered examinations to be a strong reason for not carrying out diagnostic, improvement related responses to assessment; because exams primarily control and evaluate teachers, a rational response appears to be conforming to the goal of high exam scores, rather than spending time on taking responsibility for addressing identified learning needs. This is clearly an area of much professional development for both prospective and practicing teachers in China; though, the extent to which such work will be effective will depend much on reducing the systemic impact and importance of examinations.

Note that the relationship of Exams to Irrelevant was more-or-less the same as the weak negative correlation seen between Accountability and Irrelevance among practicing teachers (Brown et al., 2011). Thus, both prospective and practicing teachers share the view that exams are only a little bit irrelevant; after all they determine and control so much of what teachers do, so they can hardly be irrelevant! The path from Exam to Life Character reinforces the Confucian notion that doing well on examinations is an expression of moral and personal virtue; a better person achieves more (Li, 2009; Pong & Chow, 2002; Tsui & Wong, 2009).

Most importantly this study provides a baseline for investigating whether the beliefs of Chinese prospective teachers can change as they transition into practice. While survey studies with practicing teachers have suggested that they do espouse beliefs about excellence and assessment that support the curricular goals of wide, holistic development, this study suggests that students who have been schooled in the last decade have not adopted matching beliefs. It is likely that there is much talk among practicing teachers about the aims and ambitions of the curriculum, but practice still focuses on maximizing examination scores. Changing educational practice so that there is alignment between curricular goals and schooling practices clearly will have to address and challenge the power of formal examinations in China. There is some evidence from Hong Kong (Carless, 2011) that teachers can use summative assessments formatively, and much stronger evidence from New Zealand (Brown & Hattie, 2012; Hattie & Brown, 2008) that standardized educational tests can be used by teachers to diagnose and improve teaching.

Nevertheless, this study has reinforced the understanding that experience in the practice of teaching is necessary for the development of beliefs about teaching and assessment that are in accordance with policy goals. However, as long as policy has both soft and hard options (Kennedy, Chan, & Fok, 2011), it is highly likely that the beliefs exhibited by prospective teachers will dominate teacher thinking in China.

References

- Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports* 94: 1015-1024.
- Brown, G. T. L., & Hattie, J. A. (2012). The benefits of regular standardized assessment in childhood education: Guiding improved instruction and learning. In *Contemporary debates in child development and education*, ed. S. Suggate and Reese, E., 287-292.
 London: Routledge.
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology* 15, no 1: 75-89. doi:10.5209/rev_SJOP.2012.v15.n1.37286
- Brown, G. T. L., Hui, S. K. F., Yu, F. 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, no 5-6: 307-320. doi:10.1016/j.ijer.2011.10.003
- Chen, J. (2007). Teachers' conceptions of excellent teaching in middle schools in the north of China. *Asia Pacific Education Review* 8, no 2: 288-297.
- Chen, J. (2010). Teachers' conceptions of excellent teaching in Chinese middle schools. PhD dissertation, the University of Auckland, New Zealand.

- Chen, J., Brown, G.T.L. Hattie, J.A.C., & Millward, P. (2012). Teachers' conceptions of excellent teaching and its relationships to self-reported teaching practices. *Teaching and Teacher Education* 28, no 7: 936-947.Bandalos, D.L., and Finney, S.J. 2010. Factor analysis: Exploratory and confirmatory. In *The reviewer's guide to quantitative methods in the social sciences*, ed. G. R. Hancock and R. O. Mueller, 93-114. New York: Routledge.
- Berry, R. 2011 Assessment reforms around the world. In *Assessment reform in education: Policy and practice*, ed. R. Berry and B. Adamson, 89-102. Dordrecht, NL: Springer.
- Bond, M.H., and Hwang, K.K. 1986. The social psychology of Chinese people. In *The psychology of Chinese people*, ed. M. H. Bond, 213-266. Oxford, UK: Oxford University Press.
- Carless, D. 2011. From testing to productive student learning: Implementing formative assessment in Confucian-Heritage settings. London: Routledge.
- Cheung, G.W., and Rensvold, R.B. 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling* 9, no 2: 233-255.
- China Civilisation Centre. 2007. *China: Five thousand years of history and civilization*. Hong Kong: City University of Hong Kong Press.

- Hattie, J. A., & 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, no 2: 189-201.
- Cortazzi, M., and Jin, L. 1996. English teaching and learning in China. *Language Teaching* 29, no 2: 61-80.
- Costello, A.B., and Osborne, J.W. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research and Evaluation 10, no* 7. http://www.pareonline.net/pdf/v10n17.pdf.
- Deng, C., and Carless, D.R. 2010. Examination preparation or effective teaching: Conflicting priorities in the implementation of a pedagogic innovation. *Language Assessment Quarterly* 7, no 4: 285-302.
- Ding, X. 2010. *Educational Reform and Development in Shanghai*. Paper commissioned by the Shanghai Academy of Education Research, Shanghai, China.
- Fan, X., and Sivo, S.A. 2007. Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research* 42, no 3: 509-529.

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 *APA Educational Psychology Handbook: Individual Differences and Cultural and Contextual Factors*, ed. K. R. Harris, Graham, S., and Urdan, T., 471-499. Washington, DC: APA.

- Feng, D. 2006. China's recent curriculum reform: progress and problems. *Planning and Changing* 27, no 12: 131-144.
- Gao, L., and Kennedy, K.J. 2011. Teachers' conception of assessment in mainland China: Categories, characteristics and implications [中国内地教师的考评观:类型、特点与后示]. Journal of Educational Studies [教育学报] 7, no 1: 39-47.
- Gao, L., and Watkins, D.A. 2002. Conceptions of teaching held by school science teachers in P.R. China: identification and cross-cultural comparisons. *Science Education* 24, no 1: 61-79.
- Han, M., and Yang, X. 2001. Educational Assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy and Practice* 8, no 1: 5-10.
- He, Y., Levin, B.B., and Li, Y. 2011. Comparing the content and sources of the pedagogical beliefs of Chinese and American pre-service teachers. *Journal of Education for Teaching: International research and pedagogy* 37, no 2: 155-171.
- Hill, M., Gilmore, A., Smith, L., and Cowie, B. 2012. *The impact of a dedicated assessment course on developing teacher education students' assessment capability*. Paper presented at the biannual meeting of the SIG Assessment and Evaluation, European Association for Research in Learning Instruction, 29-31 August, in Brussels, BE.

- Hughes, I., and Yuan. L. 2005. The status of action research in the People's Republic of China. *Action Research* 3, no 4: 383-402.
- IBM. 2011. AMOS (Version 20.0.0 (Build 817)). Meadville, PA: Amos Development Corporation.
- Kennedy, K.J., Chan, J.K.S., and Fok, P.K. 2011. Holding policy-makers to account: exploring 'soft' and 'hard' policy and the implications for curriculum reform. *London Review of Education* 9, no 1: 41-54.
- Lam, T.C.M., and Klockars, A.J. 1982. Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement* 19, no 4: 317-322.
- Li, J. 2009. Learning to self-perfect: Chinese beliefs about learning. In *Revisiting the Chinese learner: Changing contexts, changing education*, ed. C. K. K. Chan and N. Rao, 35-69.
 Hong Kong: Springer/The University of Hong Kong, Comparative Education Research Centre.
- Lin, Q., and Xun, Y. 2001. The institutional and policy development of teacher education in China. *Asia-Pacific Journal of Teacher Education and Development* 4, no 2: 5-23.
- Little, R.J.A., and Rubin, D.B. 2002. *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

- Liu, P., and Qi, C. 2005. Reform in the curriculum of Basic Education in the People's Republic of China: Pedagogy, application, and learners. *International Journal of Educational Reform* 14: 35-44.
- Louie, K. 1984. Salvaging Confucian education (1949-1983). *Comparative Education* 20, no 1: 27-38.
- Marsh, H.W., Hau, K.T., and 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* 1, no 3: 320-341.
- Min, W. 1997. Basic education and national development: Current issues and strategies for China. *Educational Research Journal* 12, no 2: 142-151.Ministry of Education. 2005. *The guidance on reforming the high school and university*
- entrance examination in experimental zone in basic education curriculum reform. Beijing, PRC: Ministry of Education.
- OECD. 2011. Strong performers and successful reformers in education: Lessons from PISA for the United States. Paris, FR: OECD Publishing.
- Pajares, M.F. 1992. Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research* 62: 307-332.
- Pong, W.Y., and Chow, J.C.S. 2002. On the pedagogy of examinations in Hong Kong. *Teaching and Teacher Education* 18, no 2: 139-149.

Rieskamp, J., and Reimer, T. 2007. Ecological rationality. In Encyclopedia of Social

Psychology, ed. R.F. Baumeister and K.D. Vohs, 273-275. Thousand Oaks, CA: Sage.

- Smith, L., Cowie, B., Gilmore, A., and Hill, M. 2012. Student and teacher educator perceptions of becoming assessment-capable teachers. Paper presented at the biannual meeting of the SIG Assessment and Evaluation, European Association for Research in Learning Instruction, 29-31 August, in Brussels, BE.
- Tsui, A.B.M., and Wong, J.L.N. 2009. In search of a third space: Teacher development in mainland China. In *Revisiting the Chinese learner: Changing contexts, changing education,* ed. C. K. K. Chan and N. Rao, 281-311. Hong Kong: Springer/The University of Hong Kong, Comparative Education Research Centre.
- Wang, G. 1996. Educational assessment in China. Assessment in Education: Principles, Policy and Practice 3, no 1: 75-88.
- Watkins, D.A., and Biggs, J.B. 2001. *Teaching the Chinese learner: Psychological and pedagogical perspectives*. The University of Hong Kong, Hong Kong: Comparative Education Research Centre and Australian Council of Educational Research.
- Watkins, D.A., and Zhang, Q. 2006. The good teacher: A cross-cultural perspective. In *Effective schools*, ed. D. McInerney, Dowson, M., and Van Etten, S., 185-204.
 Creanwich, CT: Information Aca Publishing.

Xie, A. 2001. The development and reform of teacher education in China: Theoretical issues.

Asia-Pacific Journal of Teacher Education and Development 4, no 2: 41-59.

Figure Caption

Figure 1. Structural model result showing paths from TE-TCET to TE-TCoA-C

Note. Path values are standardized beta regression weights.