



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. L., & Wang, Z. (2016). Understanding Chinese university student conceptions of assessment: Cultural similarities and jurisdictional differences between Hong Kong and China. *Social Psychology of Education*, 19(1), 151-173. doi: [10.1007/s11218-015-9322-x](https://doi.org/10.1007/s11218-015-9322-x)

Copyright

The final publication is available at Springer via <http://link.springer.com/article/10.1007/s11218-015-9322-x>

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.

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

UNDERSTANDING CHINESE UNIVERSITY STUDENT CONCEPTIONS OF ASSESSMENT:
CULTURAL SIMILARITIES AND JURISDICTIONAL DIFFERENCES BETWEEN HONG KONG
AND CHINA

Gavin T. L. Brown

The University of Auckland

Zhenlin Wang

The Hong Kong Institute of Education

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

Acknowledgement: The authors thank their research assistant, Ho Yee Tang, who carried out the validation studies, which were funded by an Internal Research Grant (#RG42/09–10) to the first author from the Hong Kong Institute of Education. Specific thanks go to colleagues who facilitated the data collection, including Dr Lamei WANG (Institute of Psychology, Chinese Academy of Sciences), Prof Lingbiao GAO (South China Normal University), Dr Yang WU (Zhengzhou Normal University), Dr Xiaorui Huang (East China Normal University), Dr Junjun CHEN (Hong Kong Institute of Education), and Prof Ming GAO (Shannxi Normal University).

Recommended Citation:

Brown, G. T. L., & Wang, Z. (2015). Understanding Chinese university student conceptions of assessment: Cultural similarities and jurisdictional differences between Hong Kong and China. *Social Psychology of Education*. Advance online publication. doi: 10.1007/s11218-015-9322-x

Abstract

Hong Kong and the People's Republic of China, while sharing historic cultural roots, have different policies for and practices of educational assessment. Student conceptions of assessment function to guide individual behaviour in response to the functions, purposes, and consequences of assessments. A new self-report questionnaire was developed to account for attitudes and beliefs detected in qualitative and pilot survey studies. In a two-group confirmatory factor analysis, an eight-factor solution, in which seven factors were dependent on a higher order factor (i.e., School Quality), was found with good fit. The seven factors were named: Societal uses, Class benefits, Accuracy, Negative Aspects, Teacher Use, and Competition. Invariance testing showed that regression weights were not equivalent between the PRC and HK students, though they were among PRC pre-degree and postgraduate students. There were statistically significant differences in factor mean scores between HK and PRC groups. Conventional Chinese cultural norms were less explanatory of results than the effect of institutional assessment policies and practices in each jurisdiction.

Keywords: Assessment; Attitudes and Beliefs; Chinese; Higher Education; Students; Survey Research

1.0 Introduction

The achievements and attitudes, especially towards teaching and learning, of Chinese learners have been widely studied and praised, especially in light of the performance of Chinese students on international comparative tests such as PISA (Chan and Rao, 2009; OECD, 2011; Watkins and Biggs, 1996). Chinese learners are perceived as being highly motivated, persistent, and effortful in their studying behaviours from a very early age. While there is some criticism that Chinese learners, perhaps because of the teaching and assessment regime, rely overly on surface, reproduction-oriented learning strategies, there is no doubt the general performance of Chinese students is meritorious.

Part of the reason for this success lies in the value given in Chinese contexts to academic achievement. An individual's merit, worth, and value are ascribed through academic performance (China Civilisation Centre, 2007); that is, a good person is one who scores well because examination results reflect the quality of the individual. Confucian thought is considered the origin of the notion in Chinese cultures (Li, 2009; Pong and Chow, 2002; Tsui and Wong, 2009) that a strong love of learning and academic success perfects the person morally and socially. Consistent with Confucian ideas, Chinese learners tend to exhibit relatively low self-efficacy due to modesty, humility, and high standards, but nevertheless through effort and persistence attain high achievement (Hau and Ho, 2008; Salili, Lai, and Leung, 2004).

Chinese parents, especially, expect students to become better academically, attitudinally, and behaviourally through schooling (Gao and Watkins, 2001) and will enforce such expectations with harsh authoritarian parenting practices (Ho, 1986; Paul, 2011; Salili, 2001). Chinese societies are generally considered to be collectivist in that the self is embedded in a network of significant people around the self (Martin, Yu, and Hau, 2014). This means that performance on assessment and academic achievement for Chinese students is tied up with their social obligations to their families (Yu and Yang, 1994). Indeed, duty to learn out of obligation to others (e.g., I learn because it is expected by family and society) and duty to work hard (e.g., If a task is difficult, I must concentrate and keep trying) were positive predictors of achievement among 152 Asian (73% of whom were from Confucian-Heritage ethnicities) university students in a New Zealand research intensive university (Peterson, Brown, and Hamilton, 2014).

Additionally, within China itself and the various Chinese dominated societies (e.g., Singapore, Hong Kong, and Taiwan) there is a long history of using examinations and tests to select and reward talent (China Civilisation Centre, 2007) and to regard high academic performance on high-stakes

examinations as a legitimate, meritocratic basis for upward social mobility regardless of social background (Cheung, 2008; Lee, 1996). Students' implicit beliefs about educational activities like assessment are expected to arise from the socialisation of their experiences within families and within the educational system they experience. Hence, if conceptions are ecologically rational (Rieskamp and Reimer, 2007), changes in uses in different contexts (i.e., political jurisdiction) may result in differences in belief systems. Simultaneously, given the strong similarities among students of Chinese ethnicity and culture around the goals of learning and academic achievement, it could legitimately be expected that they would have very similar views as to the purposes, functions, and nature of assessment.

Hence, the goal of this study was to compare and contrast responses of university students in China (PRC) and Hong Kong (HK) concerning the purposes and functions of assessment. This study allows us to examine the impact of cultural similarities (i.e., being Chinese) and policy differences between two ethnically similar populations (i.e., HK is 95% Han Chinese, while PRC is 92% Han).

1.1 Conceptions of Assessment

The beliefs, attitudes, experiences, and responses of university students about assessment arise from their experiences with formative and summative uses of assessment both prior to and within higher education itself (Ecclestone and Pryor, 2003). Beliefs act to guide, frame, and filter the actions of actors within a society or ecology (Fives and Buehl, 2012). Hence, an understanding of student conceptions of assessment in different jurisdictions, but with shared ethnicity and culture, would help us understand whether the commonalities of Chinese uses of assessment explain how university students perceive assessment. Assessments and examinations are used throughout and prior to higher education for evaluative (i.e., summative) and improvement (i.e., formative) purposes. Summative (e.g., grading, selection, or certification) and formative (i.e., feedback independent of evaluation) goals create tensions in how any assessment event or task could be understood. In other words, what might be intended by the teacher as formative (i.e., giving feedback on how to improve) may be perceived and understood by the student as highly evaluative (i.e., constructive criticism is seen as a bad result). How students understand the nature of assessment and its functions or purposes seems to shape their behavioural responses to assessment practices (Brown, 2011). For example, if students agree that assessment evaluates them, their test performance and effort tend to increase; whereas, if they conceive that assessment evaluates schools or can be ignored, their achievement and effort decreases (Brown and Hirschfeld, 2008; Brown, Peterson, and Irving, 2009; Wise and Cotten, 2009). Additionally, it has been found that accepting the purpose for

standardised testing results in greater acceptance of the importance of doing institutional accountability testing and greater effort in doing that testing (Zilberberg, Finney, Marsh, and Anderson, 2014).

There are relatively few studies of how Chinese higher education students conceive of and experience assessment, despite the dominance of high-stakes public examinations in the educational systems of China (Feng, 1995; Han and Yang, 2001; Wang, 1996; Wang, 2006) and Hong Kong (Biggs, 1998; Choi, 1999; Yan and Chow, 2002). The coherent, consistent, and competitive environment of public examinations is assumed to define student conceptions of assessment. Since formal, summative examinations determine selection, students may be less trustful of teacher-based or even peer-based assessment practices (e.g., Gao, 2009). Hong Kong university students are very aware of the negative, controlling impact of assessment on their lives (Brown and Wang, 2013) and have serious doubts as to the validity of assessment, its accuracy, and its limited utility because of its focus only on academic content (Wang and Brown, 2014). Furthermore, HK students indicated that “achievement was an obligation one had toward one’s family in order to please, show respect to, or build reputation for the family” (Wang and Brown, 2014, p.1073). At the same time, the students accepted that assessment had legitimacy for selecting the best candidates and provided upward social mobility. There was also some acceptance that assessment could motivate learning and could evaluate the quality of teachers and schools.

Two survey studies of Chinese university students concerning assessment have been found. From a survey of over 40,000 Chinese higher education students, Guo and Shi (2014) demonstrated that students who received more feedback from faculty through the classroom assessment system had higher learning outcomes. Using the multi-dimensional (eight factors and 33 items) *Student Conceptions of Assessment* (SCoA) inventory, it was found that Hong Kong students agreed more than the PRC students that assessment was irrelevant, bad and to be ignored; otherwise, the PRC students agreed more than Hong Kong students that assessment helped socially and affectively and that it contributed to improved teaching and learning (Brown, 2013). Since performance on assessment is determined by rank more than quality of work, the incentive to be cooperative and supportive in classroom environments may be reduced; there may be an incentive not to help one’s peers. Additionally, the exposure of students in Hong Kong to western individualistic values during HK’s colonisation by the west, while PRC was relatively isolated in the 2nd half of the 20th century, may also contribute to a less supportive attitude towards classmates. However, the use of the SCoA, which was developed in a Western individualist society, may

not be sufficiently coherent with the full range of Chinese perceptions of assessment. Thus, the current study attempts to develop a self-report more consistent with the Chinese context.

1.2 Assessment within the Chinese culture

China is now the second largest economy in the world and about one-quarter of the world population is Chinese (United Nations Statistics Division, 2013). Both HK and PRC share common cultural roots in Confucian values and social contexts in which selection for higher education was competitive and determined through formal assessment processes. However, despite shared cultural and historical values there are significant differences between contemporary HK and PRC in terms of educational practice and policy which may impact on assessment experiences and conceptions. More details of the differences between these two systems, despite now being part of one country, are available in Postiglione and Leung (1992).

1.2.1 Hong Kong. Recently Hong Kong has made significant changes to its educational system (e.g., Koo, Kam, and Choi, 2003; Ngan, Lee, and Brown, 2010). For example, academic aptitude tests used at the end of Primary 6 to sort students into high school bands have been stopped, the number of high school bands has been reduced to three from five (meaning schools enrol students with a broader range of achievement), and the 13th year of schooling has been moved to university (bringing Hong Kong into alignment with American and Chinese systems of 12 years of schooling followed by four years of university to reach bachelor degrees). School-based assessment procedures now form a part of the secondary school qualification system, though teacher marks are still statistically moderated by school-wide performance on end-of-year written examinations (Qian, 2014). School self-assessment procedures and annual government monitoring of school achievement have also been implemented, though endorsement of these school evaluation systems is not universal among school leaders (Ngan et al., 2010). Hence, Hong Kong schooling, despite efforts to introduce an assessment for learning policy (Berry, 2011), is still characterised by summative testing or examination with many high-stakes decisions (especially selection) based on rank order position.

1.2.2 People's Republic of China. Similarly, the education system in PRC places great emphasis on regular high-stakes public examinations (e.g., the Entrance Examination for Senior High School and the Entrance Examination for Higher Education—*zhong kao* and *gao kao*, respectively). Nonetheless, the government of China, since at least the 1990s, has introduced assessment reforms that attempt to move evaluation systems away from transmission and memorisation of 'bookish' knowledge

for purely ranking or selection purposes towards more formative, authentic, and humanistic approaches to assessment (Han and Yang, 2001; OECD, 2011) including the introduction of integrated quality assessment which emphasises judging students' personal character (Liu and Qi, 2005; PRC Ministry of Education, 2005). These reform efforts are an extension of the curricular calls for all-round development of good character and good person attributes prevalent in PRC since the mid-1950s (Wang, 1996).

However, the degree of change in the assessment system seems limited. Demand for space in higher education institutions exceeds supply, such that only half of examined candidates are awarded a place (Davey, et al., 2007); although this represents about a quarter of the 18- to 22-year-old population (PRC Ministry of Education, 2010). Further, higher education resources are not equally distributed across all regions, which favours students in urban areas along the southern and eastern coasts (Wang, 2011). Hence, intensive cramming practices to maximise performance continues¹.

Nonetheless, consistent with official policy, entrance to university is not simply a function of performance on the *gao kao*. A number of non-academic criteria exist, including students' demonstrating right moral character, privileging students who live in the same jurisdiction as the university, students' membership in a specified minority group, or having a recommendation that permits bypassing the examination altogether (Davey, De Lian, and Higgins, 2007; Wang, 2011). Consistent with market reforms, it is now possible for families with economic resources to relocate to regions that have lower entry standards or obtain extra score status illegitimately (Wang, 2011).

Given these many variations to tested performance as routes to university entrance, there are concerns within PRC about the legitimacy of the system (Davey, et al., 2007; Wang, 2011). Specifically, the system is not seen as fair for all, the content of the examinations seems to exclude important practical skills and application of theoretical knowledge, the chance of success is not high, students are expected to dedicate too much of their lives to preparation, and there it is possible the system is not immune to corruption by teachers or officials. Additionally, tuition-free higher education has been replaced by a user-pay mechanism (costing about 60% of average per capita household disposable income) meaning that tuition costs are a substantial obstacle (Wang, 2011). These structural anomalies are not limited in PRC to education. For example, despite regulation that penalises non-payment of wages, there are persistent cases of employers refusing to pay workers' end-of-year wages (Wu, 2015).

¹ See story of an examination hothouse school 'Maotanchang High School' at: <http://www.nytimes.com/2015/01/04/magazine/inside-a-chinese-test-prep-factory.html>

To summarise, Hong Kong and PRC both have gone through assessment reforms aimed at promoting more formative assessments and reducing reliance on high-stakes public examinations, although with limited success. Unlike Hong Kong, however, entrance to university has more diverse selection criteria in PRC which makes the process itself less transparent and more vulnerable to corruption and inequality. It is this level of jurisdictional difference that is assumed to create differences in student conceptions of assessment, despite the common cultural values of both Chinese jurisdictions.

1.3 Research Hypotheses

Given the high degree of similarity between PRC and Hong Kong concerning the use of assessment (i.e., high consequence public examinations) and the strong importance of obligation and duty to family shared in all Confucian-Heritage cultures, we expected that university students in both jurisdictions would have very similar responses. This is especially true of aspects of assessment having to do with meeting family obligations, being highly competitive, and experiencing significant social benefits as a consequence of performance. However, the differences in methods of entry to university suggest that PRC university students might be fairly negative about the validity, fairness, and accuracy of assessment. Furthermore, given the power of schools to determine success, through excellent teaching or through informal processes, it should be expected that PRC students will react negatively to the relative lack of equity and equality in the PRC university entry system, producing differences to the more transparent and meritocratic system HK students experience. This means that aspects of assessment that connect to the more general nature of being Chinese should have similar regression weights between Hong Kong and PRC students, while those aspects connected to the institutional specific practices of each jurisdiction should elicit quite different responses, in accordance with observed ecological differences.

2.0 Methods

2.1 Participants

This inventory was administered on a volunteer sample of university students in PRC (13 institutions) and Hong Kong (9 institutions) obtained through convenience and snowball methods. The survey was administered in May and December 2012, using both paper (81%) and online (19%) administration. About one-fifth (21%) were male, 77% female, and 1% not given. Of the 827 valid cases (Table 1), 55% were enrolled in PRC universities and 45% in HK. PRC institutions which specialise in teacher education (called Normal Universities) provided 265 cases, while an additional 353 HK cases were enrolled in a specialist teacher education institute. This means that 75% of participants were

enrolled in teacher education institutions, although, because these institutions are multi-disciplinary it cannot be guaranteed that each participant was enrolled in teacher education. Participants indicated the type of degree they were enrolled in: that is, 23% were in pre-degree diplomas or certificates (90% in PRC), 38% in bachelor degree programs (87% in HK), and 39% in post-graduate diplomas or doctoral degrees (78% in PRC). These enrolment patterns of the two groups indicate they do not come from similar populations. The high proportion of women may be consistent with education faculty enrolments, but is certainly not consistent with overall university enrolments, which in PRC are predominantly male (Wang, 2011).

Table 1. Demographic characteristics

Country and Institute	Count
<i>People's Republic of China</i>	
Shenyang Normal University	183
China Academy of Art	164
South China Normal University	55
East China Normal University	20
Guangdong University of Business Studies	17
Taizhou University	6
Foshan University	3
Southwest University	3
Beijing Normal University	2
Hangzhou Women's College	1
Northeast Normal University	1
Nanjing Normal University	1
China Pharmaceutical University	1
<i>Sub-total</i>	<i>457</i>
<i>Hong Kong</i>	
HK Institute of Education	353
Polytechnic University	6
City University	3
University of HK	3
Chinese University of HK	1
HK Vocational Training College	1
HK Baptist University	1
Lingnan University	1
Open University	1
<i>Sub-total</i>	<i>370</i>
Grand Total	827

2.2 Instrument

Large-scale studies with the *Student Conceptions of Assessment* (SCoA) inventory into student perceptions of or beliefs about the purpose of assessment have grown out of New Zealand high school

studies (Brown, 2011). Research with the SCoA has shown that students who endorse certain self-regulatory beliefs about assessment achieve better; these include, the idea that assessment (a) guides students into appropriate learning, (b) helps teachers know how to teach better, and (c) holds students personally responsible for assessment results (Brown and Hirschfeld, 2008; Brown, Peterson, and Irving, 2009). Studies with the SCoA have been carried out with higher education students in Brazil, Hong Kong, and China (Brown, 2013), with invariance between New Zealand, Hong Kong, and China students being established for parts of the inventory; suggesting that some items and scales could be used effectively in Chinese contexts.

However, qualitative research with Chinese students has suggested strongly that a new multi-dimensional, inventory was required (Brown and Wang, 2013; Wang and Brown, 2014). The qualitative studies identified 18 possible uses for or effects of assessment. These constructs included: (1) assessment is negative; (2) assessment requires effortful modesty; (3) students have validity concerns about assessment; (4) assessment has high-stakes consequences; (5) assessment focuses only on academic content; (6) assessment is lifelong; (7) assessment success permits social mobility; (8) assessment is examinations; (9) assessment success comes from ‘gaming’ strategies; (10) assessment selects; (11) success on assessment meets family obligations; (12) assessment presses and monitors students to fit society’s expectations; (13) assessment is useful for teachers and schools; (14) assessment helps personal improvement, self-reflection, and motivation; (15) assessment generates positive emotions; (16) assessment generates negative emotions; (17) assessment has too much control which needs to be escaped; and (18) assessment indicates my personal value and worth. Empirical validation of the items drafted for these 18 constructs was conducted and reported briefly.

2.2.1 Instrument Validation: Content and Language

A total of 209 items were subjected to two validity tests, based on procedures taken from Gable and Wolf (1993), before being piloted with university students. First, after splitting the items into three equally sized sets, three panels of five judges assigned items to one of the 18 construct definitions and indicated how confident their assignment was. Items for which fewer than three judges agreed on a category or which had an average confidence rating of less than two out of four were removed from the next round. Items which a majority of judges assigned, with strong confidence, to a different category than expected by the authors were retained, on the possibility that the developers had erred. Simultaneously, the functional equivalence of English and Chinese expressions for each item was

evaluated by having three panels of five bilingual Hong Kong adults rate the equivalence of items. Items which did not receive at least an average equivalence of greater than 'kind of close' from three or more judges were removed from further analysis. These processes reduced the item pool to 137 items and 17 constructs (i.e., category 17 and category 12 were merged).

2.2.2 Pilot Surveys

A pilot survey among HK and PRC university students with the 137 items was conducted by splitting the item pool into two forms (both tested in Mandarin and Cantonese) with approximately equal number of items to mitigate against fatigue. A convenience sample of 231 HK and 216 PRC students completed Part A, while 239 HK and 227 PRC students completed Part B. Each part was analysed with confirmatory factor analysis to test the validity of the intended factors. In both parts, the intended factor structure failed and a series of exploratory techniques (e.g., deletion of items with strong cross-loadings or modification indices to other factors; merging or deletion of factors which had negative error variances; conversion of a correlation between factors to a subordination of one factor to another if correlation values were $r > .95$) were used to identify a valid, acceptably fitting model. Multigroup confirmatory factor analysis was used to determine the degree of invariance between HK and PRC samples.

Part A resulted in a model with marginal fit ($k = 37$; $\chi^2 = 1760.894$; $df = 619$; $\chi^2/df = 2.845$, CFI = .81; gamma hat = .88; RMSEA = .064; SRMR = .0734). The model had two second order correlated factors (i.e., Effects and responses and School assessment) with six and three first order factors respectively. Effects and Responses contained (1) High Stakes Consequences, (2) Effortful Modesty, (3) Personal Worth, (4) Social Mobility, (5) Lifelong, (6) Validity Concerns; while School Assessment contained (7) Examination, (8) Academic Content Only, and (9) Negative. This model was not configurally invariant between HK and PRC samples. In contrast, Part B resulted in a model with good fit ($k = 37$; $\chi^2 = 2428.15$; $df = 1246$; $\chi^2/df = 1.95$, CFI = .83; gamma hat = .94; RMSEA = .045; SRMR = .063) and full invariance between HK and PRC samples. The model had a second order factor (Select and Society) consisting of seven items and which predicted six more factors (i.e., Gaming Strategies; Family Obligation; Useful for Teachers; Improvement; Positive Emotions; and Negative Emotions). Analysis of Part A and B resulted in 74 items related to 16 factors which were merged into the C-SCoA inventory reported in this paper.

2.2.3 C-SCoA inventory

The C-SCoA inventory (74 items, 16 factors) was administered in two versions; Cantonese in traditional script in Hong Kong and Mandarin in simplified script in PRC. Sample items for each factor are provided:

Selection & Societal (7 items): Assessment is used to select the best people for job and education opportunities.

Gaming Strategies (4 items): Assessment is a game which needs to be played strategically.

Family Obligation (6 items): It is my obligation to my parents to do well on assessments.

Useful for Teachers (5 items): Assessment provides information on how well schools are doing.

Improvement (6 items): I use assessments to take responsibility for my next learning steps.

Positive Emotion (5 items): When we do assessments, there is a good atmosphere in our class.

Negative Emotion (3 items): Even when I have studied, I feel I cannot control my assessment results.

High Stakes Consequences (5 items): Assessments determine my fate and future.

Effortful Modesty (3 items): No matter how hard I try, I will never be as good as others on assessments.

Personal Worth (5 items): My grades determine my value and worth to my family and society in general.

Social Mobility (4 items): Higher social status comes from good academic performance.

Lifelong (4 items): I have been assessed my whole life.

Validity Concerns (4 items): Assessment results are sufficiently accurate

Examination (4 items): School/class assessments are training for final/public examinations.

Academic Content Only (4 items): Assessments only focus on book learning and knowledge.

Negative Evaluation (4 items): Assessment is value-less.

Participants responded by indicating their degree of agreement with each statement using an ordered six-point, positively-packed, agreement scale, having two negative and four positive options. Positive-packing (i.e., increased number of response options in the positive direction) is deemed appropriate when participants are inclined to be agreeable (Brown, 2004; Klockars and Yamagishi, 1988; Lam and Klockars, 1982); a situation likely for academically successful students enrolled in higher education. Furthermore, good measurement characteristics (i.e., ordered response options with relatively equal intervals) were found with this rating scale in a study of Chinese higher education students (Deneen, Brown, Bond, and Shroff, 2013).

2.3 Analysis

All analyses were conducted with AMOS version 20 software (IBM, 2011). First, data were cleaned by removing cases with more than 10% missing responses and imputing missing values with the expectation maximisation procedure. Second, confirmatory factor analysis was used to determine if the data fit the empirically and theoretically defined 16 factors. Third, due to poor or inadmissible fit, exploratory factor analysis, following procedures described in Courtney (2013), was used. 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 more conceptually-sound solution being preferred. Fourth, the EFA model was further trimmed in AMOS by removing items with weak loadings ($\lambda \leq .40$), with strong cross-loadings ($\lambda \geq .30$), or with strong modification indices to other factors (sum MI > 30).

Although the inventory elicits responses using a six-point, ordinal agreement scale, maximum likelihood estimation was used since scales of this length can be treated as continuous variables (Finney and DiStefano, 2006). Fit of a confirmatory model is determined by inspection of a number of indices (Fan and Sivo, 2005; Hu and Bentler, 1998). Current standards suggest that models do not need to be rejected if the root mean square error of approximation (RMSEA) is <.08, the standardised root mean residual (SRMR) is <.08, the comparative fit index (CFI) and gamma hat indices are >.90 (Fan and Sivo, 2007; Yu, 2002).

Multi-group confirmatory factor analysis with sequential nested invariance testing was used to evaluate the degree to which the model was invariant across groups (Cheung and Rensvold, 2002). This procedure sequentially tests the statistical equivalence of a model between groups by examining the impact of constraining a set of parameters to be equal upon the comparative fit index (CFI) measure of fit. Invariance testing generally follows the sequence: (1) configural equivalence of paths; (2) equivalent regression weights for factors to items (metric); (3) equivalent intercepts of items at factors (scalar); (4) equivalent structural paths from factors to factors; and (5) equivalent covariances among factors. Equivalent residuals for items and factors are not required to make comparisons. Configural invariance is accepted if RMSEA < .05, while equivalence for all other tests is accepted if the $\Delta CFI < .01$.

3.0 Results

We report first the factor analytic statistical modelling that best represents Chinese higher education student conceptions of assessment responses. Then we evaluate the equivalence of the model across the jurisdictions and conclude with an analysis of the mean score differences.

3.1 Model Determination

After removing mis-fitting items, an inter-correlated model of eight factors was found ($N=861$; $k=33$; $\chi^2=20134.12$, $df=467$; $\chi^2/df=4.36$, $p=.04$; CFI=.84; gamma hat=.91; SRMR=.056; RMSEA=.061, 90%CI=.058--.064). The factors discovered were named: Societal uses, Class Benefits, School Quality, Accuracy, Negative Aspects, Teacher Use, Family Effects, and Competition (Table 2 provides items contributing to each factor). We considered that the common cultural factors for Confucian-Heritage societies were: Competition, Societal Use, Exam Accuracy, and Family Effects. These factors point to the Chinese use of high-stakes examinations to award on a competitive basis societal benefits, motivated by obligation or duty to the family. In contrast, we considered that the factors sensitive to jurisdictional differences in institutional practices and policies as: Teacher Use, School Quality, Class Benefit, and Negative Effects. These factors point to the different applications of assessment by teachers and students within school contexts and the different bases for judging school quality seen in each jurisdiction. Finally, the negative effects factor points to the different reasons within each jurisdiction that invalidate the legitimacy of assessment.

Table 2. C-SCoA factors, items, and standardised loadings.

#	Factor & Statements	Loading		
		HK B	PRC Pre	PRC Post
<i>Culturally-similar factors due to shared Confucian heritage</i>				
F8 Family Effect				
51	My family values me only if I do really well on assessments ^{ns}	.71	.69	.69
41	My grades determine my value and worth to my family and society in general ^{ns}	.65	.69	.69
64	I am smart only if I get 100% or am best in class ^{ns}	.57	.62	.62
60	A high rank in school is how I please my family**	.76	.65	.65
46	My family’s reputation depends on my performance on assessments***	.77	.65	.65
71	My family is very much affected by my assessment results***	.77	.60	.60
F4 Accuracy				
27	Assessment results can be depended on ^{ns}	.62	.60	.60
54	Assessment results are sufficiently accurate**	.77	.66	.66
F1 Societal uses				
7	Higher social status comes from good academic performance. ^{ns}	.65	.58	.56
70	Success in society depends on doing well on assessments**	.70	.58	.58
56	With higher grades, I can gain a better position in society***	.87	.63	.63
37	Good grades lead to a better career path***	.84	.66	.66
F7 Competition				
2	Assessments have a huge impact on my place in society ^{ns}	.51	.49	.49
1	Assessment is used to select the best people for job and education opportunities ^{ns}	.50	.47	.47
30	Assessment is a competition to select the best***	.80	.66	.66

#	Factor & Statements	Loading		
		HK B	PRC Pre	PRC Post
6	Assessment never stops; it's part of being alive***	.40	.65	.65
<i>Institutional Policy Factors Unique to Jurisdiction</i>				
F2 Class benefits				
39	When we do assessments, there is a good atmosphere in our class***	.43	.63	.63
50	Our class becomes more supportive when we are assessed**	.79	.71	.71
65	Assessment encourages my class to work together and help each other*	.83	.77	.77
42	Assessment makes our class cooperate more with each other ^{ns}	.79	.75	.75
F3 School Quality				
24	Assessment prepares students for examinations***	.56	.35	.35
63	Assessment provides information on how well schools are doing*	.51	.62	.62
33	Growing up means you get assessed regularly ^{ns}	.53	.49	.49
31	Assessment measures the worth or quality of schools ^{ns}	.59	.66	.66
F5 Negative Aspects				
18	Assessment is limited to what can be learned in books***.	.81	.66	.66
4	Assessments only focus on book learning and knowledge***	.73	.57	.57
19	I am useless if I don't get top in my class***	.44	.68	.68
26	My classmates and peers are better at assessments than I am***	.31	.55	.55
25	I ignore or throw away my assessment results***	.22	.52	.52
36	Assessment results ignore all the other things I can do well**	.53	.36	.36
14	Assessment results are filed & ignored ^{ns}	.43	.52	.52
F6 Teacher Use				
53	Assessment helps teachers track my progress***	.56	.71	.71
73	Teachers use my assessment results to see what they need to teach me next**	.69	.57	.57

Note. HK = Hong Kong; PRC = People's Republic of China; B = Bachelor's degree; Pre = Pre-degree

diploma or certificate; Post = Post-graduate diploma or degree; all values are standardised; statistical

significance of loading differences between HK and PRC: ns=not statistically significant, * $p < .05$,

** $p < .01$, *** $p < .001$.

3.2 Model Equivalence Evaluation

Two-group analysis resulted in a positive not definite covariance matrix among the PRC participants, suggesting over-factoring had occurred for that group. Inspection of the factor inter-correlations for the whole group (Table 3) suggested that the School Quality factor could be modelled as a source factor, predicting all other factors; this produced an admissible solution with acceptable fit, when run as a two-group model ($N_{\text{PRC}} = 458$, $N_{\text{HK}} = 370$; $k = 33 * 2$ groups; $\chi^2 = 3052.32$, $df = 976$; $\chi^2/df = 3.13$, $p = .08$; CFI = .78; gamma hat=.93; SRMR=.086; RMSEA=.051, 90%CI=.049--.053). After constraining

the factor to item regression weights (metric) to be equivalent, the model was found not to be invariant ($\Delta\text{CFI} = .010$).

Table 3. C-SCoA factor inter-correlations

Factor	Pearson Inter-correlations							
	1	2	3	4	5	6	7	8
1. Teacher Use	—							
2. Competition	.53	—						
3. School quality	.69	.88	—					
4. Class benefits	.62	.28	.40	—				
5. Societal uses	.50	.77	.80	.28	—			
6. Accuracy	.76	.59	.72	.61	.57	—		
7. Negative Effects	.45	.34	.47	-.04	.49	.27	—	
8. Family Effects	.27	.51	.62	.21	.60	.47	.57	—

Note. Values > .70 in bold; values for F3 in red; $N = 861$.

However, since the distribution of students by country was different by the level of study, invariance testing was also carried out on three groups (i.e., PRC pre-degree $n = 170$, HK bachelor degree $n = 272$, and PRC post-degree diplomas and degrees $n = 251$). This model had good fit, with configural invariance ($N=693$; $k=33 * 3$ groups; $\chi^2 = 3836.11$, $df = 1570$; $\chi^2/df = 2.44$, $p = .12$; CFI = .72; gamma hat = .94; SRMR = .112; RMSEA = .046, 90%CI = .044--.048), and although metric equivalence was demonstrated ($\Delta\text{CFI} = .009$), this was not sufficient to make inter-group comparisons. All two-way comparisons among these three groups were conducted and strict equivalence was found for PRC pre-degree vs. PRC post-degree groups (i.e., metric $\Delta\text{CFI} = .006$; scalar $\Delta\text{CFI} = .001$; structural weights $\Delta\text{CFI} = .001$; structural covariances $\Delta\text{CFI} = .000$; structural residuals $\Delta\text{CFI} = .003$; and item residuals $\Delta\text{CFI} = .007$), while metric equivalence was not found between either PRC group and the HK group ($\Delta\text{CFI} = .012$, and .011, PRC pre-degree and PRC post-degree respectively). Hence, it is concluded that responses within PRC, despite differences in degree type, were statistically equivalent, and PRC students in general, regardless of degree status, had statistically different responses to those of HK students. The HK group responses indicated they were sampled from a separate population to the PRC students and this is more likely to be due to differences in jurisdiction than degree level.

We had expected that factors related to more general Confucian-Heritage conceptions of assessment would be statistically equivalent between groups. In contrast, we expected the jurisdictionally specific factors to elicit non-equivalent responding. As Table 2 shows, seven of the 16 items in the Confucian-Heritage factors had non-statistically significant differences in loading, while 13 of the 17 of the jurisdictionally specific items had statistically significant differences in loading. This means that 44% of the items in the Confucian-Heritage group matched expectations, while 76% of jurisdictionally specific

items matched expectations. Nonetheless, it is reasonably clear that conventional Chinese cultural values are not strong predictors of the similarity of responses between PRC and Hong Kong. It does seem, in contrast, that jurisdictional or institutional factors do contribute more to understanding the differences in responses between PRC and Hong Kong.

3.3 Factor Means Evaluation

Although the invariance testing indicated that the groups had responded in non-equivalent ways to the items, it is informative to examine the factor means to help appreciate the nature of the differences between groups. Table 4 shows the factor means (created by averaging responses to each factor's contributing items) and MANOVA (main effects and interaction for country [PRC vs. HK] and level [PRC pre, HK B, PRC Post] and effect size comparisons between the three groups. For seven factors, except Class Benefit, the HK Bachelor degree group had the highest means, while the PRC Pre-degree group had the lowest means. Except for Class Benefit and Family Effects, the scale of differences between those two groups was statistically significant, with moderate to large effect sizes. The HK Bachelor and PRC Post-degree group means were much closer, with only half reaching statistical significance (i.e., Societal Use, $F = 55.05, p < .001, R^2 = .09$; School Quality, $F = 9.94, p .002, R^2 = .02$; Negative Aspects, $F = 25.25, p < .001, R^2 = .04$; and Competition, $F = 4.43, p = .04, R^2 = <.01$). Tukey post-hoc HSD test indicated that for three factors the HK Bachelor mean was distinguishable from the two PRC groups (i.e., Societal Use, School Quality, Negative Aspects); for three factors the PRC Pre-degree group was distinguishable from the PRC Post-degree and HK Bachelor groups which were indistinguishable (i.e., Accuracy, Teacher Use, and Competition); while, as per the ANOVA and effect size results, two factors produced indistinguishable means across the three groups (Class Benefit and Family Effects).

Within the HK Bachelor group, four factors had means >3.50 (i.e., close to or greater than moderately agree); these had to do with the external uses of assessment to evaluate schools, generate societal benefits with success, generate competition, and inform teachers. Among the two PRC groups Teacher Use had a similar mean score, as did Competition for the PRC Post-degree group. It is noteworthy that only two factors had means below 3.00 (i.e., slightly agree) in the HK Bachelor group, whereas the PRC Pre group had five and the PRC Post group had four means below the same threshold. This indicates a more general unwillingness to agree with these statements among PRC respondents. The least endorsed belief had to do with impact of assessment on family relationships and the negative aspects

of assessment. Compared to previous studies, this result suggested Chinese students, more so Hong Kong Bachelor degree students, viewed assessment predominantly in terms of external purposes, even rejecting the idea that assessment was inaccurate, while not associating assessment performance with negative familial consequences.

Table 4. C-SCoA(HE) scale means by degree group with comparisons

C-SCoA(HE) Scale	<u>Group Descriptive Statistics</u>						<u>Difference Statistics</u>		
	<u>HK B</u>		<u>PRC Pre</u>		<u>PRC Post</u>		<u>MANOVA</u>		<u>Effect size</u>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i> ₍₂₎	<i>p</i>	Cohen's <i>d</i>
<i>Culturally-similar factors</i>									
Competition	3.86	0.89	3.40	0.88	3.69	0.97	13.02	<.001	.50
Societal Use	3.65	1.09	2.84	0.86	2.98	0.95	45.64	<.001	.82
Exam Accuracy	3.01	0.87	2.54	0.93	2.91	1.00	13.88	<.001	.50
Family Effects	2.49	0.97	2.43	0.90	2.49	0.93	0.27	.76	.06
<i>Jurisdictional Policy Factors</i>									
Teacher Use	3.98	0.89	3.50	1.05	3.83	1.07	12.57	<.001	.49
School Quality	3.63	0.81	2.54	0.89	3.39	0.88	17.35	<.001	.56
Class Benefit	3.00	0.90	3.09	0.94	3.09	1.03	0.65	.52	.09
Negative Effects	2.80	0.75	2.39	0.75	2.46	0.82	19.13	<.001	.52

Note. HK = Hong Kong; PRC = People's Republic of China; B = Bachelor's degree; Pre = Pre-degree diploma or certificate; Post = Post-graduate diploma or degree; $N_{(HK\ B)} = 272$; $N_{(PRC\ Pre)} = 170$; $N_{(PRC\ Post)} = 251$.

4.0 Discussion

4.1 Summary of Findings

A configurally equivalent model of Chinese university students' beliefs about assessment was found between students in Hong Kong and the People's Republic. The statistical model indicates that students do discriminate among these competing purposes for assessment, but see them as facets of evaluating schools. Assessment as an indicator of school quality is the central factor and all other aspects of assessment are perceived as being dependent on this core concept. However, there were statistically significant differences in item regression weights and factor mean scores between the two jurisdictions that seem to be attributable more to institutional differences in how assessment is implemented than to the

conventional assumption that shared Chinese cultural factors determine the beliefs and attitudes of Chinese students. Because this study was an anonymous survey, it was not possible to seek explanations from the participants for the current results. Hence, we offer the following speculations for the observed results.

4.2 Cultural Chinese views of Assessment

Four factors were considered to reflect the conventional cultural Chinese view of assessment being reliable, competitive examinations that produce important social consequences and fulfil obligations to family. We expected similarity between PRC and HK students especially in these factors and this was more or less not the case.

The relatively low scores for the family effects factors, equally endorsed by all three groups, indicated that students' perceived place in and value to the family was not contingent upon tested performance. This suggests that Chinese families are not seen as cold and punitive, at least, by these students, who are successfully enrolled in universities. The rejection of the idea that family reputation is dependent on student performance may be a strong divergence from the traditional notion of collectivist responsibility to ancestors and family urged by Confucian philosophy. For the PRC students, this may well be a reflection of 'single child' policies post-1979 and free-market economic practices post-1984. These two systemic changes may have contributed to greater parental warmth towards a sole child and the greater belief that educational success is no longer necessary for family well-being. For HK students, longer exposure to western individualism and a greater diversity of economic opportunities may have defused the notion that academic success is driven by familial obligations. Or much more simply, it may be that having successfully entered universities, these students had already fulfilled familial obligations and so no great negative effect would be incurred as long as the students completed their studies successfully.

Nonetheless, three items in this factor had statistically significant differences in regression weights that were stronger in Hong Kong. These three items speak strongly to the classic Confucian notion of familial piety or obligation. It is possible these items elicit stronger association for HK students, even though weaker level of endorsement, because of the anti-Confucian movement in 1973 (*pī Lín pī Kǒng*) as part of the PRC Cultural Revolution (1966-1976). In contrast, Hong Kong students gave higher factor means and stronger item loadings (four out of six items) for the factors of assessment accuracy and societal uses.

We suspect that within Hong Kong there is a stronger basis for attributing societal success to assessment results and that examinations are reliable. With agencies such as the Independent Commission against Corruption and with a free press to conduct public scrutiny of government agencies such as the Hong Kong Examination and Assessment Authority, there is a reasonably strong basis for Hong Kong students to have confidence in the reliability of the examination system and the meritocratic application of examination results for university entrance. This social fabric stands in contrast to the multi-faceted basis for entry to university seen in the PRC system and the relative difficulty in ensuring transparent and consistent application of the legal system in the PRC. The alternative (and potentially corrupt) methods of university entrance in PRC may also explain why the competition factor had a higher mean among HK students and why Item #30 had a stronger loading among HK students—it is clear that HK university students have this experience (Brown and Wang, 2013; Wang and Brown, 2014).

We have, as yet, no explanation for why one item (#6) had a stronger loading for the PRC students than the HK students. Perhaps this statement captures the Chinese expression 考考考，老师的法宝；分分分，学生的命根 [exam, exam, exam, teacher's magic weapon, grade, grade, grade, students' lifeblood] reinforcing the widespread perception of the perpetual nature of examinations and assessment in PRC.

4.3 Jurisdictional Effects on Views of Assessment

Four factors were considered to reflect the differences in institutional policies and practices in the two different legal jurisdictions. These have to do with assessment being a positive classroom influence, a valid measure of school quality, a tool for teachers, and being negative. As pointed out earlier most of these items had statistically different loadings and three of the factors had large differences in mean scores. While we had not expected differences, these factors do potentially reflect jurisdictional differences and thus, under the principle of ecological rationality, differences are observed in these samples.

While the School Quality factor was the central predictor of all other factors for both groups, the mean score for this factor was considerably higher for HK students. The centrality of this factor may reflect a somewhat simplistic logic (i.e., good schools have high results; high results equal quality). The association of school quality and high examination scores (item #24) might be strengthened by the systemic nature of school segregation, especially in HK where schools are identified by bands, in contrast to the PRC where a multiplicity of status factors (e.g., money, connections, and location) can be used to

gain entrance. Nonetheless, a greater loading on item #63 by PRC students might suggest that the public comparison of university entrance results by schools has an impact on their perceptions of how assessments reflect school quality. Fundamentally, everyone knows which schools are good by the public dissemination of their examination results. Since university access depends heavily upon tested performance, schools that have higher mean scores on those measures are considered to be high quality. Perhaps a form of reflected glory is at work in which attendance at a successful school ensures examination success for the individual.

Belief that teachers use assessment to improve teaching was higher for HK and the postgraduate students of PRC and the loadings of these items differed by statistically significant margins. The notion that teachers track students with assessment loaded on the factor more for PRC students, while teachers diagnosing next teaching steps loaded more strongly for HK students. This latter result may reflect the established use of school-based assessments in HK secondary schools which makes more transparent to students how teachers adjust teaching in response to tested performance (Carless, 2011; Gao, 2009). In contrast, conventional practice in PRC involves drilling for examinations and monitoring student performance regularly (Chen, 2015).

While the differences in mean score for class benefits were trivial and not statistically significant, there were statistically significant differences in the item loadings. Two class cooperation items received stronger loadings in HK than PRC, while good atmosphere had stronger loading in PRC than HK. This is consistent with HK higher education practices that assess students in and through cooperative, collaborative group work (e.g., Auyeung, 2004; Jackson, 2005; Li, 2001).

The negative aspects factor elicited quite different responses from HK and PRC students. There was a moderate mean score difference between jurisdictions, with stronger endorsement in HK than PRC. Three items (Items #18, 4, and 36) to do with the narrow focus of assessment on book learning had stronger loadings in HK, while two items to do with rank order comparison (Items #19, 26) had stronger loadings in PRC. Likewise, an item to do with ignoring assessment results (#25) had stronger loadings in PRC. It seems HK students are more attuned to the artificial nature of examined knowledge and skill and seem to long for a broader and more valid of learning (Brown and Wang, 2013; Wang and Brown, 2014). The greater sensitivity to normative comparisons combined with greater ignoring of results in PRC may reflect the public display of examination results (i.e., results are listed in rank order on noticeboards within institutions for inspection by all); in other words, everyone can see my results, so it is important

that I outperform my classmates. Since the PRC students gave this factor a very low score (i.e., they disagreed with this factor), it would seem that the strong loading on this item indicates a rejection of the habit of ignoring the results. This means that for PRC students, much more than HK students, high rank matters, while HK students instead reject the validity of examinations.

4.4 Limitations of study

Notwithstanding the statistical significance and the plausibility of the results reported here, there are a few limitations. As a natural consequence of voluntary access and participation restrictions the sampling, albeit large, is not systematically representative and equivalent across the three groups studied. Likewise, the oversampling from three institutions and from teacher education institutions limits the generalisability of the results. Hence, further replication studies which ensure greater variety and equivalence of disciplines, degree levels, and institutions are required. Additionally, the speculations we have given to explain the results are testable research topics in future studies using techniques such as think-aloud or retrospective stimulation, or by specifying the type or function of assessment which the participant must evaluate. Future research that links this inventory to measures of achievement emotions, learning approaches and conceptions, and causal attributions would also give validity to the speculations we have offered.

4.4 Significance and Conclusions

Nonetheless, this study contributes to our understanding of the role social and cultural factors play on the beliefs, attitudes, and values students have. This study demonstrates that the Chinese-Student Conceptions of Assessment (Higher Education) inventory is sensitive enough to pick up nuanced differences between HK and China students' individual beliefs and emotional reactions to assessment, while at the same time successfully identifying culturally invariant factors that speak to the Confucius heritage of the Chinese people. This study suggests that sharing Chinese identity is not enough to understand how university students in Chinese contexts understand and respond to assessment. Instead, student conceptions of assessment seem to reflect more the ecological factors of the educational environment in which they are schooled. This study should help university educators and administrators to better understand the diversity in Chinese students' attitudes and behaviour in terms of learning processes and outcomes.

This study helps further establish that environmental differences can invalidate comparative studies since responses to good inventories are often not equivalent across jurisdictions, and highlights the

importance of developing culturally appropriate measurement tools incorporating both a theory driven and bottom up empirical perspective. Students' conceptions of assessment reflecting an agentive response to the micro and macro environments might be sensitive to ongoing educational and assessment reforms. An evolving historical perspective guarantees ecological validity and will require constant updating of research inventories.

References

- Auyeung, L. H. (2004). Building a collaborative online learning community: A case study in Hong Kong. *Journal of Educational Computing Research*, 31(2), 119-136.
- Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Policy, Principles and Practice*, 18(2), 199-211.
- Biggs, J. (1998). The assessment scene in Hong Kong. In P. Stimpson & P. Morris (Eds.), *Curriculum and assessment for Hong Kong: Two components, one system* (pp. 315-324). Hong Kong: Open University of Hong Kong Press.
- Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, 94(3), 1015-1024.
doi: 10.2466/pr0.94.3.1015-1024
- Brown, G. T. L. (2011). Self-regulation of assessment beliefs and attitudes: A review of the Students' Conceptions of Assessment inventory. *Educational Psychology*, 31(6), 731-748. doi: 10.1080/01443410.2011.599836
- Brown, G. T. L. (2013). Student conceptions of assessment across cultural and contextual differences: University student perspectives of assessment from Brazil, China, Hong Kong, and New Zealand. In G. A. D. Liem & A. B. I. Bernardo (Eds.), *Advancing Cross-cultural Perspectives on Educational Psychology: A Festschrift for Dennis McInerney* (pp. 143-167). Charlotte, NC: Information Age Publishing.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice*, 15(1), 3-17. doi: 10.1080/09695940701876003
- Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Beliefs that make a difference: Adaptive and maladaptive self-regulation in students' conceptions of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.) *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 159-186). Charlotte, NC: Information Age Publishing.
- Brown, G. T. L., & Wang, Z. (2013). Illustrating assessment: How Hong Kong university students conceive of the purposes of assessment. *Studies in Higher Education*, 38(7), 1037-1057. doi: 10.1080/03075079.2011.616955

- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-Heritage settings*. London: Routledge.
- Chan, C. K. K., & Rao, N. (Eds.). (2009). *Revisiting the Chinese learner: Changing contexts, changing education*. Hong Kong: Springer/The University of Hong Kong, Comparative Education Research Centre.
- Chen, J. (2015). Teachers' conceptions of approaches to teaching: A Chinese perspective. *The Asia-Pacific Education Researcher*, 24(2), 341-351. doi: 10.1007/s40299-014-0184-3
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Cheung, T. K.-Y. (2008). An assessment blueprint in curriculum reform. *Journal of Quality School Education*, 5, 23-37.
- China Civilisation Centre. (2007). *China: Five thousand years of history and civilization*. Hong Kong: City University of Hong Kong Press.
- Choi, C.-C. (1999). Public Examinations in Hong Kong. *Assessment in Education: Principles, Policy & Practice*, 6(3), 405-417.
- 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), 1-14.
- Davey, G., De Lian, C., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education*, 31(4), 385-396. doi: 10.1080/03098770701625761
- Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2013). Telling the difference: A first evaluation of an outcome-based learning innovation in teacher education. *Asia-Pacific Journal of Teacher Education*, 41(4), 441-456. doi: 10.1080/1359866X.2013.787392
- Ecclestone, K., & Pryor, J. (2003). 'Learning careers' or 'Assessment careers'? The impact of assessment systems on learning. *British Educational Research Journal*, 29(4), 471-488.
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, 12(3), 343-367.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509-529.

- Feng, Y. (1995). From the imperial examination to the national college entrance examination: The dynamics of political centralism in China's educational enterprise. *Journal of Contemporary China*, 4(8), 28-56.
- 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.
- Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. (2nd ed.). Boston, MA: Kluwer Academic Publishers.
- Gao, M. (2009). Students' voices in school-based assessment of Hong Kong: A case study. In D. M. McInerney, G. T. L. Brown & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 107-130). Charlotte, NC: Information Age Publishing.
- Gao, L., & Watkins, D. A. (2001). Towards a model of teaching conceptions of Chinese secondary school teachers of physics. In D. A. Watkins & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 27-45). Hong Kong: University of Hong Kong, Comparative Education Research Centre.
- Guo, F., & Shi, J. (2014). The relationship between classroom assessment and undergraduates' learning within Chinese higher education system. *Studies in Higher Education*, (ahead-of-print), 1-22. doi: 10.1080/03075079.2014.942274
- Han, M., & Yang, X. (2001). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice*, 8(1), 5-10.
- Hau, K. T., & Ho, I. T. (2008). Editorial: Insights from research on Asian students' achievement motivation. *International Journal of Psychology*, 43(5), 865-869.

- Ho, D. Y. F. (1986). Chinese patterns of socialization: A critical review. In M. H. Bond (Ed.), *The psychology of the Chinese people* (pp. 1-37). Hong Kong: The Chinese University of Hong Kong.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.
- IBM. (2011). Amos 20 [computer program] (Version Build 817). Meadville, PA: Amos Development Corporation.
- Jackson, J. (2005). An inter-university, cross-disciplinary analysis of business education: Perceptions of business faculty in Hong Kong. *English for Specific Purposes*, 24(3), 293-306. doi: <http://dx.doi.org/10.1016/j.esp.2004.02.004>
- Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement*, 25(2), 85-96.
- Koo, R. D., Kam, M. C., & Choi, B. C. (2003). Education and schooling in Hong Kong under one country, two systems. *Childhood Education*, 79(3), 137-144.
- 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.
- Lee, W. O. (1996). The cultural context for Chinese learners: Conceptions of learning in the Confucian tradition. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 25-42). Hong Kong: Comparative Education Research Centre and Australian Council for Educational Research.
- Li, J. (2009). Learning to self-perfect: Chinese beliefs about learning. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 35-69). Hong Kong: Springer/The University of Hong Kong, Comparative Education Research Centre.
- Li, L. K. Y. (2001). Some Refinements on Peer Assessment of Group Projects. *Assessment & Evaluation in Higher Education*, 26(1), 5-18. doi: 10.1080/0260293002002255
- Liu, P., & 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(1), 35-44.
- Martin, A. J., Yu, K., & Hau, K. T. (2014). Motivation and engagement in the 'Asian Century': a comparison of Chinese students in Australia, Hong Kong, and Mainland China. *Educational Psychology*, 34(4), 417-439.

- Ngan, M. Y., Lee, J. C. K., & Brown, G. T. L. (2010). Hong Kong principals' perceptions on changes in evaluation and assessment policies: They're not for learning. *Asian Journal of Educational Research and Synergy*, 2(1), 36-46.
- OECD. (2011). *Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*. Paris, FR: OECD Publishing.
- Paul, A. M. (2011). The roar of the Tiger Mom. *Time (Asia)*, 177(4), 24-30.
- Peterson, E. R., Brown, G. T. L., & Hamilton, R. J. (2013). Cultural differences in tertiary students' conceptions of learning as a duty and student achievement. *The International Journal of Quantitative Research in Education*, 1(2), 167-181. doi: 10.1504/IJQRE.2013.056462
- Pong, W. Y., & Chow, J. C. S. (2002). On the pedagogy of examinations in Hong Kong. *Teaching & Teacher Education*, 18(2), 139-149.
- Postiglione, G. A., & Leung, J. Y. (1992). *Education and Society in Hong Kong. Toward One Country and Two Systems*. Hong Kong, SAR: Hong Kong University Press.
- PRC 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.
- PRC Ministry of Education. (2010). All levels of education gross enrollment rate. In Educational Statistics in 2009, Ministry of Education, Government of China. Retrieved from <http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/s4959/201012/113470.html> (accessed September 8, 2011).
- Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice*, 21(3), 251-270. doi: 10.1080/0969594X.2014.915207
- 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.
- Salili, F. (2001). Teacher-student interaction: Attributional implications and effectiveness of teachers' evaluative feedback. In D. A. Watkins & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 77-98). Hong Kong: University of Hong Kong, Comparative Education Research Centre.

- Salili, F., Lai, M. K., & Leung, S. S. K. (2004). The consequences of pressure on adolescent students to perform well in school. *Hong Kong Journal of Paediatrics*, 9(4), 329-336.
- Tsui, A. B. M., & Wong, J. L. N. (2009). In search of a third space: Teacher development in mainland China. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 281-311). Hong Kong: Springer/The University of Hong Kong, Comparative Education Research Centre.
- United Nations Statistics Division (December 2013). *GDP and its breakdown at current prices in US Dollars*. Retrieved from <http://unstats.un.org/unsd/snaama/dnlist.asp>.
- Wang, G. (1996). Educational assessment in China. *Assessment in Education: Principles, Policy & Practice*, 3(1), 75-88.
- Wang, L. (2011). Social exclusion and inequality in higher education in China: A capability perspective. *International Journal of Educational Development*, 31(3), 277-286. doi: 10.1016/j.ijedudev.2010.08.002
- Wang, X. B. (2006). *An introduction to the system and culture of the College Entrance Examination of China* (RN-28). New York: The College Board, Office of Research and Analysis.
- Wang, Z., & Brown, G. T. L. (2014). Hong Kong tertiary students' conceptions of assessment of academic ability. *Higher Education Research & Development*, 33 (5), 1063-1077. doi: 10.1080/07294360.2014.890565
- Watkins, D. A., & Biggs, J. B. (Eds.). (1996). *The Chinese learner: Cultural, psychological and contextual influences*. Hong Kong: Comparative Education Research Centre.
- Wise, S. L., & Cotten, M. R. (2009). Test-taking effort and score validity: The influence of the students' conceptions of assessment. In D. M. McInerney, G. T. L. Brown & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 187-205). Charlotte, NC: Information Age Publishing.
- Wu, X. (2015, January 24-25). Make it easy for workers to get pay on time. *China Daily*, page 5.
- Yan, P. W., & Chow, J. C. S. (2002). On the pedagogy of examinations in Hong Kong. *Teaching and Teacher Education*, 18(3), 139-149.
- Yu, A. B., & Yang, K. S. (1994). The nature of achievement motivation in collectivist societies. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi, & G Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 239-250). Thousand Oaks, CA: Sage.

- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. (Unpublished doctoral dissertation), University of California, Los Angeles, Los Angeles, CA.
- Zilberberg, A., Finney, S. J., Marsh, K. R., & Anderson, R. D. (2014). The role of students' attitudes and test-taking motivation on the validity of college institutional accountability tests: A path analytic model. *International Journal of Testing*, 14(4), 360-384. doi: 10.1080/15305058.2014.928301

Biographical Notes

Associate Professor Gavin T L Brown, PhD, is the Director of the Quantitative Data Analysis and Research unit in the Faculty of Education and Social Work at the University of Auckland, New Zealand. His research focuses on cross-cultural study of educational assessment and its impact on teachers and students.

Assistant Professor Zhenlin Wang, PhD, is in the Department of Psychological Studies in the Faculty of Education and Human Development at the Hong Kong Institute of Education. Her primary research focuses on children's theory of mind development, and its implication in children's social functioning, as well as early childhood teaching and learning.