Understanding outcome-based education changes in teacher education: Evaluation of a new instrument with preliminary findings

Christopher Deneen
Research Assistant Professor
The University of Hong Kong, Hong Kong
852 9177 7108
deneen@hku.hk
The University of Hong Kong Faculty of Education Pok Fu Lam Road Pok Fu Lam, Hong Kong, SAR

This paper and the research contained herein meets all relevant ethical guidelines, including adherence to the legal requirements of the study country, Hong Kong, SAR, China and the ethical guidelines and procedures of the associated institution, the Hong Kong Institute of Education. Ethical approval was granted by the relevant institutional ethics committee. Participants were informed prior to their participation in the research and gave informed consent. Approval has been granted by the relevant ethics committee for the institutional name to be publicly revealed and such revelation is deemed important to the reported context of the study.

Keywords: Student evaluation, teacher education, Rasch analysis, factor analysis

Secondary Authors:

Gavin T. L. Brown
Associate Professor
The University of Auckland, New Zealand
Trevor G. Bond
Professor
James Cook University, Queensland, Australia
Ronnie Shroff
Senior Teaching and Learning Officer
Hong Kong Baptist University, Hong Kong

Acknowledgements ***(Optional)***

The authors would like to thank Dr. Lixun Wang for his invaluable help in implementing this study.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Understanding outcome-based education changes in teacher education: Evaluation of a new instrument with preliminary findings

Outcomes Based Education (OBE) is a current initiative in Hong Kong universities, with widespread backing by governments and standards bodies. However, study of students' perceptions of OBE and validation of understanding these perceptions are lacking. This paper reports on the validation of an OBE-specific instrument and resulting preliminary findings.

Instrument responses (*n*=89) were analysed using Rasch and exploratory/confirmatory factor analyses. Both approaches identified two dimensions (i.e., evaluation and comparison). Challenges and modifications to the instrument items and their relationships to constructs are discussed. Preliminary findings suggest students did not perceive significant differences between OBE and traditionally-organised courses. Lack of explicit discussion of OBE with the students may have denied students the ability to make fully informed evaluations of OBE innovations.

Implications for instrument validation and evaluation of initiatives in an OBE context are discussed as is the broader issue of transparency in teacher education curriculum design and implementation.

Keywords: Outcome-based, student evaluation, instrument validation, teacher education, mixed methodology, Rasch analysis, factor analysis

Introduction

Higher education institutions and specifically teacher education programs constantly seek to improve the quality of their teaching and graduate outcomes. Current trends in higher education quality assurance focus on first identifying the intended outcomes or goals of a course or program and then aligning teaching, learning, and assessment so as to maximise the likelihood that students achieve those outcomes or goals (Baron & Boschee, 1996; Deneen, 2009; Spady, 1994; Webb, 2009). This stands in contrast to more traditional approaches of understanding program quality and success by focusing on inputs, such as student recruitment demographics, graduate placement, and staff acquired research funding (Nusche, 2007), and understanding course success through faculty input (e.g., teaching) as opposed to student results and outcomes (Barr & Tagg, 1995). This shift from an input and instructor-centric model to a student outcome model has been characterized as a significant and fundamental paradigm shift in how tertiary institutions are conceptualizing program and course quality

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

(Barr & Tagg, 1997). An outcomes approach focuses and interconnects the components of an educational system around the anticipated achievement of learners (Biggs & Tang, 2007). In such a model, learning experiences are "designed backwards" (Wiggins & McTighe, 2005) from the anticipated learning outcomes, so that they may synchronize with and support the achievement of outcomes. In combining a focus on the learner with determinable results, outcomes-based educational approaches aim to inform a systematic, learner-centric approach to quality enhancement and assurance (Webb, 2009).

One important question in attempting to change from an inputs-based model of curriculum design to an outcomes-based model is whether such changes make a difference to students' educational experience. More generally, attention to student perspectives of the quality of higher education is a major objective of student evaluation of teaching and course evaluation (Richardson, 2005). Given the opportunity, students are capable of accurately evaluating valuable educational experiences. This is inclusive of what might be considered good teaching (Greimel-Fuhrmann & Geyer, 2003; Kember & Wong, 2000) and correlations between student rating of teachers and their academic achievement have been positive (Watkins, Marsh, & Young, 1987). Students' recognition of instructor accomplishment may extend beyond didactic engagement into more innovative practices that are in alignment with outcome based education (OBE) changes, including active student engagement (Irving, 2004). Further, students' evaluation of course constructs such as group interactions and assessment correlate positively to academic achievement and has been shown to have high validity and reliability (Richardson, 2005). It is, therefore plausible and desirable that higher education students should be able to provide insights into multiple characteristics of an outcome-based learning framework.

This paper reports findings from the first phase of an outcome-based innovation within one higher education institute in Hong Kong. The research is designed to address two areas critical to early-stage innovation. Specifically, this research seeks to:

1. confirm the properties of a survey instrument designed specifically to explore an outcomes model of course implementation
2. report preliminary findings regarding students' course perceptions

This paper examines how learners perceived a course designed according to outcome-based principles through a newly developed student self-report inventory, the Student Evaluation of Outcome-Based Learning Survey (SEOBLS version 1). There are a number of practical reasons for developing an instrument to evaluate perceptions of outcome-based learning (OBL) from a student perspective. Firstly, because one of the key aims of OBL is to ensure that all learners achieve high quality outcomes expressed through cognitively deeper and richer learning experiences, a special-purpose instrument could provide data about students' higher-order cognitions. Secondly, feedback from students on perceived effectiveness of OBL initiatives should allow improvements in the design and

implementation of OBL course design. Thirdly, development of an OBL-focused student evaluation instrument may be of benefit beyond the Hong Kong Institute of Education (HKIEd) to researchers and programme developers seeking to consider the use of OBL in higher education, and specifically teacher education.

As with all courses, the OBE trial course has been evaluated using HKIEd's standardised student evaluation of teaching survey (SET). The SET is general in approach with some items bearing relationship to those of the SEOBLS version 1 instrument (ex. SET- *What was taught matched the aims and objectives of the course outline*; SEOBLS- *The stated learning outcomes agree with what is actually taught in the course*). However, upon careful review, the SET items were deemed not to match the intended OBE design changes closely enough to serve as a robust basis for evaluating the innovation. OBE design changes represent a significant change from traditional input models (Barr & Tagg, 1997); therefore, we deemed it beneficial to produce an instrument focused specifically on the intended areas of change, emerging from an OBE design perspective.

This study aimed to identify how learner perspective into higher education course delivery and design might inform the process of introducing similarly designed courses into teacher education programs. Insights into the newly developed inventory are discussed, as well. Finally, the study also discusses how the application of two different methods of analysis (i.e., factor analysis and Rasch analysis) produced similar as well as distinct results.

Context

The University Grants Committee (UGC) of Hong Kong launched in 2005 an outcome-based education (OBE) campaign to promote OBE at the eight funded institutions within Hong Kong. The stated goals of this initiative were improved and enhanced student learning and teaching quality (Ewell, 2005; Stone, 2005).

The reported pilot study was carried out at the Hong Kong Institute of Education, a UGC-funded tertiary institution focused on teacher education and complementary social sciences and humanities disciplines. The Institute has 403 academic staff and 4830 full-time equivalent students in three faculties (i.e., Arts & Sciences, Education Studies, and Languages) and grants bachelors, masters, and doctorate degrees in education. It is by far the largest teacher training institution in the region, accounting for 84% of Hong Kong primary teachers, 80% of trained kindergarten teachers and 30% of all secondary school teachers.

The pilot study took place during the fall 2009 semester in an English Language Department course (ENG 1244, Introduction to Language Studies). This 12-week Bachelor of Education course has the purpose of providing "an introduction to the main fields of linguistic enquiry, furnishing an initial knowledge base in the areas of language, linguistics and communication" (Wang, 2009, p 1). ENG 1244 is a mandatory course offered early in

students' degree programme designed to "provide a coherent overview of the interrelationships amongst ensuing major academic content modules on linguistics" (Wang, 2009, p. 1).

One element that made this an attractive choice for a survey test-bed is the absence of electives at this point in the program. Students were at this point, taking the same courses. In addition, the courses students had taken up to this point had not undergone any OBE-specific enhancement nor had they been part of a parallel change affecting teaching and learning, outcomes, or assessment. Courses that students were asked to compare to were at the time of the survey deployment following a traditional, input-oriented model, similar to that described in Table 2.

As part of the OBE trial, course objectives were converted to Course Intended Learning Outcomes (CILOs) (see Table 1). A notable change is the shift in each statement from the act of enabling (i.e. enable students to…) to an expectation of demonstration by the student (i.e. *Upon successful completion of the course, students will be able to:)*, exemplifying the paradigm shift from focus on instructors to students, and a shift in focus from inputs to outcomes.

Table 1. Comparing Course Objectives with CILOs

Traditional course outline Course Objectives	**OBL course outline** Course Intended Learning Outcomes (CILOs)
To enable students to:	Upon successful completion of the course, students will be able to:
1. demonstrate an understanding of the sub-domains of linguistics, enquiry, furnishing an initial morphology and semantics, discourse, sociolinguistics and psycholinguistics;	CILO1. analyse and articulate the nature, structures and functions of English language as a rich and complex system;
2. demonstrate an understanding of issues in each sub-domain pertinent to education, in particular, to the teaching and learning of language;	CILO2. apply principles of language to the specifics of the English language system;
3. demonstrate an ability to analyze and discuss core aspects of language, linguistics and communication;	CILO3. demonstrate a clear understanding of the roles and value of different varieties of English and their uses;
4. develop competence in academic reading skills.	CILO4. demonstrate high level of English academic literacy in speaking,

Traditional course outline Course Objectives	**OBL course outline** Course Intended Learning Outcomes (CILOs)
	writing and online contexts CILO5. work collaboratively in an effective way to develop English academic literacy and subject knowledge. (Communication Skill & Social Interaction Skill)

(Modified from Wang, 2010)

Teaching and learning activities were aligned to the five CILOS using three criteria: (1) the course must build on task-relevant knowledge; (2) the learner must be relevantly active; and (3) as learning progresses, students must engage reflectively and gain a clear overview of the learning engagement (Biggs and Tang, 2007). Assessment tasks and grading rubrics were designed using Killen's (2007) principles for alignment of assessment tasks. As visible in Table 2, accompanying this shift to outcomes and enabling pedagogy is a substantial shift in course assessment as well. An explicit relationship is evidenced between the assessment and the specific CILOs that assessment is linked to. Assessment tasks have been diversified, allowing students increased and varied opportunities for demonstrating outcome achievement.

In order for assessments to have impact on student learning, they were designed to:

(1) be aligned with the intended outcomes,
(2) be focused on knowledge and skills that are valued in the disciplinary context and beyond,
(3) have content validity with the domain of the course,
(4) provide valuable feedback, and
(5) be reliable and fair.

Table 2. Comparison of assessment tasks in the old course outline and in the new OBL course outline

Old course outline Two assessment tasks	**New OBL course outline** Four assessment tasks
1. An individual written essay (1000 words) on a module related topic (**60%** of the total grade).	1. An individual written essay (1000 words) on a module related topic. (40%) : *CILO 1, 2, 3, & 4*
2. A group task in which each group member	2. A group task in which each group member

would contribute (1000 words) to a chapter of a student authored academic book based on the topics introduced in the module. Peer editing among group members will be required and members in the same group will receive the same group grade (40% of the total grade).	would contribute (1000 words) to a chapter of a student authored book based on the topics introduced in the module. Peer editing among group members will be required and members in the same group will receive the same group grade. (40%) *CILO 1, 2, 3, & 4* 3. A 15-minute group presentation of the framework of the chapter that each group will write. Members in the same group will receive the same group grade. (10%) CILO 1, 2, 3, 4, & 5 4. 10 short online weekly quizzes during the module. (10%) CILO 1, 2, & 3

Additionally, the course instructor identified four principles by which the assessment rubrics were developed:

> (1) The assessment criteria should map with the CILOs.
> (2) The criteria should be articulated in a transparent way.
> (3) The criteria need to be observable and easy to be measured with evidence.
> (4) The criteria should demonstrate what our expectations on the students are.
>
> (Wang, 2010, p. 5)

Hence, the course under investigation was clearly devised to be a legitimate expression of OBE, in alignment with the previously discussed principles of OBE.

Participants

The fall 2009 ENG 1244 course consisted of 89 enrolled first-year students and three instructors. Students were a mix of approximately 70% Hong Kong native and 30% mainland Chinese, in the age range of 18-20. Students were divided into three groups with one instructor per group. All three groups had English as the mode of instruction (EMI). The SEOBLS version 1 survey was administered simultaneously across all three groups, at the end of the course.

Instrument

The SEOBLS version 1 course evaluation instrument was designed to address three areas: course intended learning outcomes, teaching & learning activities, and assessment tasks. OBE, as reviewed, involves an expansive range of outcome based educational components. However, it was deemed appropriate at this early stage of the innovation to focus on specific components of OBE most directly impacted by the OBE course modifications. Hence, the current inventory described in this study does not cover the full range of possible outcome based modifications.

Questionnaire items were also classified as to whether they focused on the propriety, feasibility, utility, and accuracy aspects of the OBE innovation (JCSEE, 1998).

Propriety has to do with proper conduct of procedures, understanding that rights and interests of all involved stakeholders have to be respected. Example: *Course methods of evaluating student work are fair and appropriate*. Feasibility is concerned with the degree to which procedures were realistic, viable, and practical, given constraints on time and resources. Example: *The amount of time I spend in this course working towards exams and graded materials is reasonable*. Utility has to do with usefulness of procedures for intended users. Example: *Tutor lectures contribute to my understanding of the course content*. Accuracy focuses on clarity, validity, and reliability of the OBE procedures. Examples include: *The stated learning outcomes are clear and understandable*; *the stated learning outcomes agree with what is actually taught in the course*. Comparative items require students to compare the OBE course with other courses they were taking. As noted, these courses utilized a traditional non-OBE format. Example: *Course workload, relative to other courses, is greater*.

As these were first year students, the survey specified a comparison of courses they had taken that semester. The survey consisted of 26 questions; questions 1-12 were related to the four identified constructs, with an even distribution of three items per construct, randomized. Questions 13-17 asked students to compare the OBE-adjusted course to other courses they had taken. Questions 18-24 solicited demographic information and the final two questions related to voluntary participation in a focus group. This option was not explored in the pilot phase due to low response.

Participants were asked to indicate how strongly they agreed or disagreed with each statement using a six-point, positively-packed agreement scale (Lam & Klockars, 1982). In a positively-packed scale, there are more agreement options (i.e., slightly, moderately, mostly, and strongly agree) than disagreement options (i.e., mostly disagree and strongly disagree). Brown (2004) has argued that a positively-packed response scale is useful when participants are likely to agree with all statements. Given evidence that positive conformity is a recognized response style within Chinese sample groups (Bond & Hwang, 1986), we

hypothesized that the greater number of options within the positive range would elicit greater variation in responses than if only two response points were used to capture positive orientation as seen in a conventional Likert scale. The Rasch analysis of the responses was used to determine the validity of the positively-packed rating scale.

Analysis

In an early evaluation such as this, the intent behind analysis may be conceptualized two ways. First, there is statistical analysis intended to identify strengths and weaknesses of the survey questionnaire itself. Second, there is the intention to understand student evaluations of initial OBE innovations. While the first intent may be seen as the major intent of a pilot, initial findings can provide valuable information for subsequent larger-scale studies. However, in order to legitimately accomplish both intentions, it is necessary that the quality characteristics of the questionnaire be established.

Two techniques were used to analyze the quality characteristics of the student survey responses: 1) Rasch analysis and 2) factor analysis. Rasch analysis is ideal for determining the extent to which items belong to a single dimension and where items sit within that dimension (Bond & Fox, 2007), while factor analysis, using both exploratory and confirmatory approaches is ideal for determining number of dimensions in a data set and which items belong to each dimension. Readers should note that these two analyses rely on dissimilar statistical frameworks; Rasch uses a logistic model to predict the probability of responding to a specific option, while factor analysis uses the variance and covariance of items with the assumption that the data are normally distributed. Hence, it is expected that dissimilar though potentially complementary results may arise. Although one or the other analytic approach is more typically used, using both together to provide an enhanced understanding of results has been used to positive effect in student evaluation instrument refinement (Richardson, 2005).

Rasch Analysis

Rasch analysis requires that scales exhibit the property of unidimensionality in order to be considered as having interval level measurement properties, including, crucially, iterative scale units (Andrich, 1988; Bond & Fox, 2007). Unidimensionality requires that empirical-actual item / person residuals are not so large as to indicate that the actual data vary significantly from the Rasch model expectations. Weighted and unweighted residual statistics are calculated as mean squares of those residuals (i.e., infit and outfit mean squares; $0.75 < x < 1.3$) or transformed into probability statistics (i.e., infit and outfit t / z; $-2.0 < x < +2.0$). Further, RFA (Rasch factor analysis of the item / person residuals) can be used to check residual patterns for evidence of second (or further) dimensions. Inspection of the category response curves is used to verify that respondents used the rating scale response options in a

meaningful, hence, measurable way. All Rasch analyses were conducted using Winsteps (Linacre, 2009).

Factor Analysis

Factor analysis is a well-established means of determining whether responses to survey items aggregate mathematically into conceptually meaningful pools (Kline, 1994). The inter-item variance/covariance matrix is examined to identify the degree to which item responses are explained by latent traits which are assumed to logically explain respondent behaviour. The pattern matrix of regressions from the latent trait to each item is used to identify items whose variance is most strongly explained by underlying constructs. Exploratory factor analysis identifies likely pools of items explained by a shared trait, while confirmatory factor analysis tests the fit of the proposed factor structure to the data by constraining items to be explained only by their respective factor (Jöreskog, 2007). Determining the number of factors is best achieved by inspection of multiple fit indices (Hoyle & Duvall, 2004) rather than reliance on the scree plot or eigenvalues >1.00 rule (Bandalos & Finney, 2010).

Robust models have values >.95 for goodness of fit indices (e.g., comparative fit index [CFI], gamma hat) and values <.05 for badness of fit indices (e.g., root mean square error of approximation [RMSEA] and standardized root mean residual [SRMR]), while goodness of fit >.90 and badness of fit <.08 are generally understood to indicate acceptable levels of fit (Bandalos & Finney, 2010). Since the χ^2 test is demonstrably overly-sensitive with large samples and complex models, Marsh, Hau, & Wen (2004) have argued that acceptable models have $p>.05$ for the ratio of χ^2 divided by degrees of freedom (i.e., χ^2/df) and this is the value we report for testing the statistical significance of χ^2 for the model. With the data set for this pilot, one missing value was imputed using the expectation maximization procedure (Dempster, Laird, & Rubin, 1977).

Results

Table 3 shows the frequency of response of each item and the mean scores and standard deviation of each item.

Statistical Characteristics of the Questionnaire: Rasch Analysis

The validity of the six-point positively packed rating scale was inspected by examining the Andrich (1978) rating-scale model analysis of the probability distributions relative to difficulty of endorsement for the six rating scale responses. This approach has common threshold response curves for all items (see Figure 1). The relatively uniform peaks

for the response categories suggested that all options were used sufficiently and indicated that the responses were not disturbed by the positively packed rating scale response options.

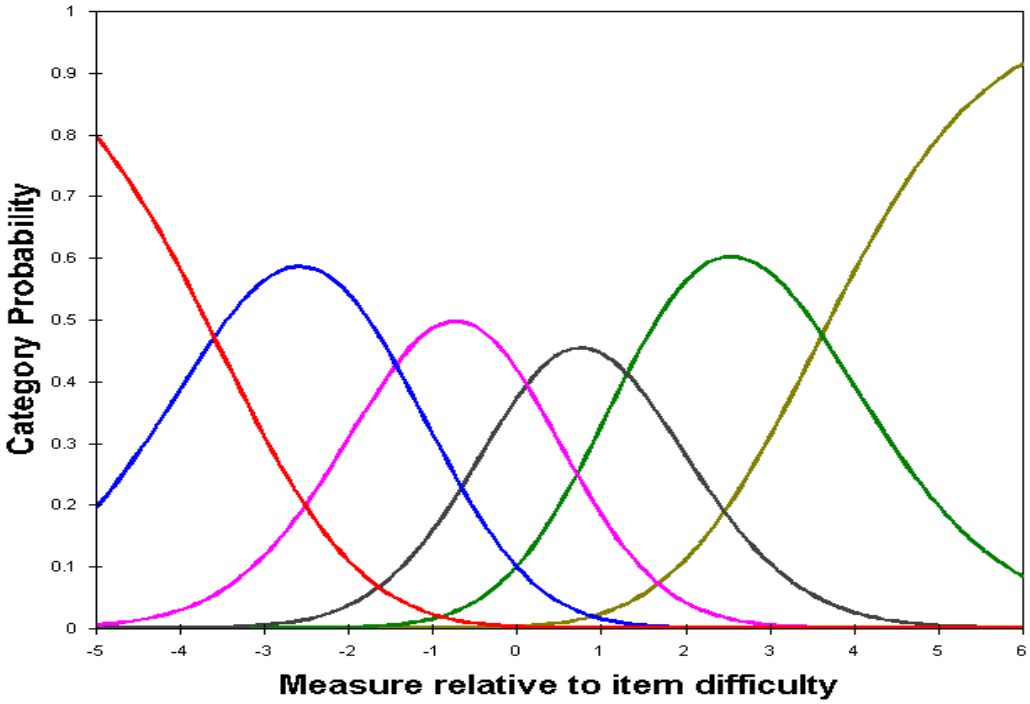

Figure 1. Rating scale model category probability curves for selected items.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Table 3: Descriptive statistics from the survey

Item		Response option frequencies						*M*	*SD*
		1 Strongly Disagree	2 Mostly Disagree	3 Slightly agree	4 Moderately Agree	5 Mostly Agree	6 Strongly Agree		
1	The stated learning outcomes of the course have a valuable relationship to my degree programme.	0	2	6	30	38	13	4.97	0.70
2	The stated learning outcomes of the course have a valuable relationship with my practice as a teacher.	0	3	14	32	31	9	4.75	0.84
3	The stated learning outcomes are clear and understandable.	1	2	8	18	42	18	5.00	0.84
4	The stated learning outcomes agree with what is actually taught in the course.	1	2	7	24	46	8	4.74	0.77
5	Tutor lectures contribute to my understanding of the course content.	1	1	10	24	39	14	4.87	0.87
6	Course activities are well prepared and carefully explained.	0	3	7	25	28	26	5.09	0.86
7	The required reading materials /texts are helpful and practical.	0	2	14	23	37	13	4.69	0.78
8	There is a clear relationship between the teaching and learning activities and the stated course outcomes.	1	2	10	29	41	6	4.78	0.66

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Item		Response option frequencies						*M*	*SD*
		1 Strongly Disagree	2 Mostly Disagree	3 Slightly agree	4 Moderately Agree	5 Mostly Agree	6 Strongly Agree		
9	Lectures, group work and other learning activities have a clear relationship to course assessment.	1	3	4	26	35	20	5.00	0.76
10	The feedback I have received on examinations/graded materials helps my improvement.	0	2	22	34	28	3	4.34	0.90
11	Course methods of evaluating student work are fair and appropriate.	1	4	16	20	44	4	4.66	0.79
12	Examinations/graded materials test the course content as emphasized by instructor.	1	1	10	29	38	10	4.53	0.80
13	The amount of time I spend in this course working towards exams and graded materials is reasonable.	1	2	11	33	37	5	4.41	0.76
14	Course difficulty, relative to other courses, is greater.	0	13	24	26	23	3	3.69	1.18
15	Course workload, relative to other courses, is greater.	0	11	25	25	23	5	4.00	1.02
16	Course pace, relative to other courses, is faster	1	19	33	19	17	0	3.84	1.02
17	My enjoyment of this course, relative to other courses is greater.	0	5	26	30	25	3	4.50	0.72

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Analyses used the single-parameter Rasch model which requires that discrimination values of the polytomous items sum to one and are statistically equivalent across items; items which deviate by more than chance are rejected as being misfitting (i.e., either being not discriminating as much as or discriminating much more than the mean). While this approach satisfies the requirements of measurement (Bond & Fox, 2007), it may, from the point of view of factor analysis, remove high quality items. Table 4 shows the item statistics in 'measure order' with the most difficult to endorse items at the top along with the degree of item fit to the single-parameter Rasch model.

Three items (i.e., #14 *Course difficulty, relative to other courses, is greater*, #15 *Course workload, relative to other courses, is greater*, and #16 *Course pace, relative to other courses, is faster*) had consistently high misfit to the underlying single dimension associated with the remaining 14 items (i.e., items 1 – 13 & 17). It should be noted that the item-score correlation for these three items was considerably weaker (i.e., in the range .30 to .33) than all of items representing the single Rasch dimension. Reverse scoring of these misfitting items did not alter the conclusion that these three items were not part of the same underlying single parameter OBE measurement construct. Inspection of these three items reveals that they have some common characteristics which may contribute to their belonging to a different dimension; the three rejected items 14-16 were the most difficult to endorse and were the explicitly comparative items (i.e., *relative to other courses*) about difficulty, workload, and pace. This might be preliminary evidence that the process of comparing the OBE course to other courses is a separate dimension from that of evaluating the OBE course *per se*.

Another two items (i.e., items #1 *The stated learning outcomes of the course have a valuable relationship to my degree programme*; and #8 *There is a clear relationship between the teaching and learning activities and the stated course outcomes*.) also had excessive outfit values and were rejected from the main dimension; misfit here is likely to be a function of overly-high discrimination values seen in r_{pb} values $\geq$.80. The two highly discriminating items were located in the relatively easy to endorse region (i.e., b values = -.52 and -.15). A case could be made under different item response theory assumptions for retaining these two items, regardless of their rejection by the Rasch model. Nonetheless, the Rasch analysis results support dropping or at least significantly modifying items and this was done, subsequent to this study. It is possible with additional items and data from a much larger sample that a different model could arise; this will be explored in a subsequent study.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Table 4. Item Rasch fit statistics for full and trimmed inventories (in item entry order).

			Full inventory							Trimmed inventory					
					Infit		Outfit					Infit		Outfit	
Item	Raw Score	*N*	Measure order	*se* (logit)	Mean Square	*z*	Mean Square	*z*	Total *r*	Measure order	Total *r*	Mean Square	*z*	Mean Square	*z*
1	410	89	-0.52	0.15	0.45	-4.6	0.48	-4.3	0.86	.03	.85	.64	-2.7	.64	-2.7
2	385	89	-0.01	0.14	0.75	-1.8	0.76	-1.7	0.75	.72	.75	1.00	0.00	.99	0.00
3	419	89	-0.72	0.15	0.97	-0.1	0.97	-0.1	0.72	-.80	.76	.97	-0.1	.76	0.77
4	400	88	-0.41	0.15	0.69	-2.3	0.7	-2.2	0.75	-.53	.78	.90	-.60	.95	-.30
5	408	89	-0.48	0.15	0.88	-0.8	0.93	-0.4	0.72	-.74	.76	1.05	.40	1.04	0.30
6	423	89	-0.81	0.15	1.02	0.2	1.03	0.2	0.76	-.03	.82	.80	-1.4	.74	-1.5
7	401	89	-0.33	0.14	0.78	-1.5	0.77	-1.6	0.76	.27	.82	.78	-1.60	.75	-1.8
8	392	89	-0.15	0.14	0.53	-3.8	0.55	-3.6	0.8	-.31	.83	0.71	-2.0	.75	-1.6
9	418	89	-0.7	0.15	0.90	-	0.90	-	0.77	-.75	.82	.81	-1.3	.83	-1.1

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

			Full inventory							Trimmed inventory					
					Infit		Outfit					Infit		Outfit	
Item	Raw Score	*N*	Measure order	*se* (logit)	Mean Square	*z*	Mean Square	*z*	Total *r*	Measure order	Total *r*	Mean Square	*z*	Mean Square	*z*
10	364	89	0.4	0.14	0.72	0.6 - 2.1	0.72	0.7 - 2.1	0.68	1.16	.68	1.20	1.30	1.18	1.20
11	381	89	0.07	0.14	1.16	1.1	1.25	1.6	0.60	.08	.62	1.54	3.00	2.26	5.20
12	399	89	-0.29	0.14	0.74	- 1.9	0.77	- 1.6	0.73	-.59	.74	1.04	0.30	1.04	0.30
13	385	89	-0.01	0.14	0.70	- 2.2	0.71	- 2.1	0.73	-.19	.75	1.04	0.30	1.00	0.00
14	335	89	0.93	0.13	1.97	5.4	2.03	5.6	0.30	—	—	—	—	—	—
15	342	89	0.8	0.13	1.96	5.3	1.97	5.3	0.33	—	—	—	—	—	—
16	299	89	1.57	0.13	1.61	3.7	1.71	4.2	0.38	—	—	—	—	—	—
17	351	89	0.64	0.14	0.87	- 0.9	0.89	- 0.7	0.66	1.69	.64	1.32	2.00	1.42	4.2
M	383.1	88.9	0.00	0.14	0.98	- 0.4	1.01	- 0.3		.00		.99	-0.2	1.05	0.1
SD	33.3	0.2	0.65	0.01	0.44	2.8	0.45	2.8		.72		.24	1.5	.39	1.9

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Over half (54%) of variance was explained by the main scale (i.e., 12 items), and almost one quarter (23.9%) of the unexplained variance could be attributed to the misfitting items (updated item parameters and fit statistics for the trimmed inventory are shown in Table 4). This suggested that two dimensions, rather than just one, were present in the data. Consequently, 12 items were retained in the main Rasch analysed scale. These items include learning objectives (3 items), teaching and learning activities (4 items), assessment practices (3 items), and one item each for time requirement and enjoyment of the course. Hence, the items appear to have formed a single evaluative scale of course experience—to what extent did you enjoy this OBE course? The mean score for this scale was 4.44 (*SD*=.74) and Cronbach's alpha estimate of reliability was α=.93. However, given the discrepancy in item and person mean values (.00 and .65 respectively) the analysis suggests that items considerably more difficult to endorse related to the course evaluation aspects of this scale were needed.

Statistical Characteristics of the Questionnaire: Factor Analysis

Exploratory factor analysis (i.e., maximum likelihood estimation with oblique minimization) identified three factors with eigen values >1.00. However, one factor had only one item, so a two-factor solution was forced in accordance with recommendations from Bandalos and Finney (2010). Factor 1 (i.e., course evaluation) had 14 items evaluating the course (i.e., items 1-13 and 17) and Factor 2 (i.e., course difficulty) had three items to do with workload, speed, and difficulty of the OBE course compared to other courses (i.e., items 14, 15, and 16). Confirmatory factor analysis had marginal fit to the two-factor solution (χ^2=241.75, *df*=118, [χ^2/*df*=2.05, *p*=.15]; CFI=.87; gamma hat =.86; RMSEA=.109; SRMR=.072). Inspection of modification indices identified three items with strong cross-loadings to other items or factors (i.e., #16, 1, and 11). By trimming these three items, a two-factor solution had acceptable fit (χ^2=117.22, *df*=76, χ^2/*df*=1.54, *p*=.21; CFI=.95; gamma hat =.94; RMSEA=.079; SRMR=.052) (Figure 2). Thus, both analytic approaches supported the notion that the questionnaire responses were based on two independent traits.

The inter-correlation between factors was weak (*r*=.11), indicating relative independence of the two factors. The mean score for Factor 1 Course Evaluation was 4.45 (*SD*=.76), which is almost identical to that of the Rasch derived Course Evaluation scale. The mean for Factor 2 Course Difficulty was 3.80 (*SD*=1.01). The difference in means had a large Cohen's (1992) effect size (*d*=.72). This means students agreed with the Course Evaluation factor considerably more than the Course

Difficulty factor. In other words, this course did everything the students expected and was not, in their estimation, too hard.

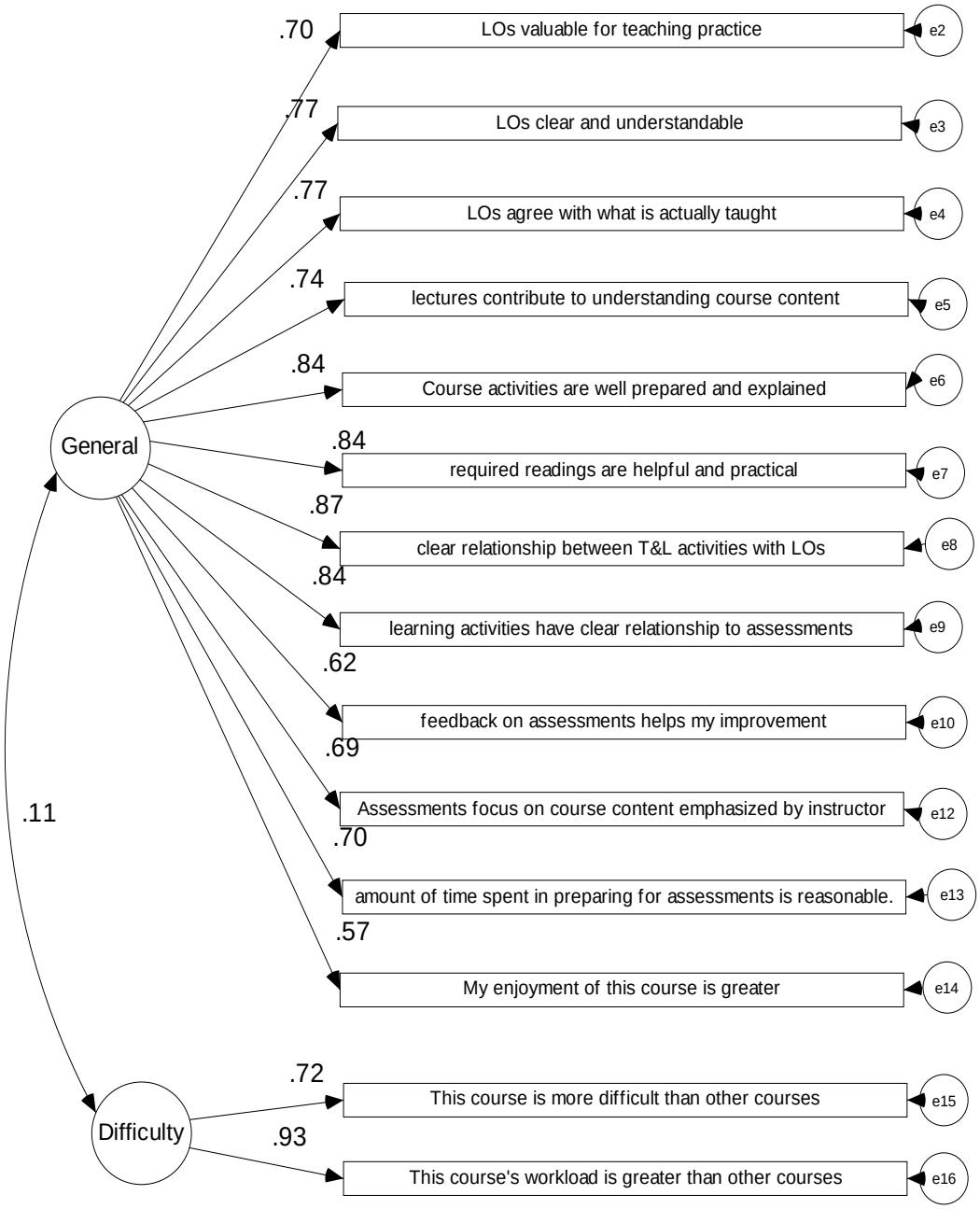

Figure 2. Two factor model of student responses to OBE course

Discussion

In response to the first intention of confirming the properties of the instrument, the two statistical analyses identified strengths and improvement needs for the SEOBLS questionnaire itself. The Rasch analysis suggested that at least three items allocated to the comparative dimension were considerably more difficult to endorse

and had much less discriminating power than did items retained in the Course Evaluation scale. This, along with both analyses, suggests considerably more items need to be created to allow better investigation into how students evaluate OBE revised courses comparatively with conventional courses. Despite proceeding from different assumptions, both Rasch and factor analyses clearly indicated that two scales were needed to meaningfully understand student evaluations of the course. Both approaches generally agreed on which items belonged to the two dimensions, so just using a summed total score would be inadequate for understanding student responses. The quality of fit for the retained items was good and evidence was found for the legitimacy of the positively packed rating scale. Future plans are underway for a larger scale study with an enhanced SEOBLS instrument for deeper inquiry into students' perceptions. It should be noted with a larger sample which has a more diverse experience of OBE that a different model and pattern of results may arise.

Despite instrument limitations, results from this study were illuminating in a number of areas related to the second intention of the pilot: to understand initial student responses to the OBE course enhancements.

The tentative conclusion reached from examining the descriptive statistics (see Table 3) in light of the two scale scores was that for these students, their experience in the OBE course was not a radical departure from a "regular" course. Students seemed, for the most part, satisfied with course quality in the areas predicted as having maximum OBE impact (i.e. teaching and learning, outcomes, and assessment), but neither did they seem to consider the course easier or harder than other courses within their experience. Hence, the principal preliminary finding of an outcome-based course innovation was that students did not perceive the OBE course as something significantly different. There are several plausible explanations for this result.

One possibility is actual implementation and/or fidelity may have been limited. This is a possibility worth considering as implementation and fidelity are relevant concerns in any professional development or change initiative. However, interviews with the instructors, presentations by the organizing instructor based on his own developmental grant research, and a comprehensive examination of course documents suggested that that implementation of and fidelity to the OBE framework was achieved. We may, therefore reasonably suggest that lack of fidelity does not account for students' limited perception of change.

Another possibility was that students were not equipped to tell the difference between this OBE designed course and other courses they took in the same semester. While actual changes to outcomes, teaching and assessment were meticulously documented, what did not emerge from the analysis was a picture that these changes were made explicit or transparent to students. The instructors did not explain to the

students that OBE was being experimented with and hence, denied them the opportunity to become directly aware of the innovation. Thus, although significant changes were made in the design and implementation of the course, in the absence of preparing students to making informed critiques of design and implementation, students may subsequently not have noted, reported, or reflected on differences in course characteristics. Students may not have been able to tell the difference because they were not made aware of the difference. This might be compounded by the sample; the participating students were in their first year of tertiary education. They were asked to compare the pilot course to other courses they were currently taking. Literature suggests students may evaluate their educational experiences differently, depending on the stage they are at in their educational career (Richardson, 2005). This is further substantiated by the difficulty of fit for the specific items calling for a comparison between the OBE course and other courses. The explanation may lie in an incongruity between what is asked of the student and what, as critical assessors, they are capable of delivering. Given the implications for exploring OBE as well as the broader target of enhancing student evaluation, it is worth exploring in further studies the degree to which students are prepared to evaluate the methods of a course, either comparatively or through criteria.

This research suggests another significant incongruity. A core goal of HKIEd is to prepare future teachers to be able to function as competent professionals in Hong Kong schools. It is a source of concern, then that students transitioning into educational professionalism were neither informed of, nor able to detect a fundamental paradigm shift in the approach to curriculum and education. Students learn to become teachers not simply through transmission of content, but also through the apprenticeship of practice and an awareness of the framework of the curriculum they engage with as students. If teacher educators do not make explicit to prospective teachers how a course has been designed, an opportunity for relevant apprenticeship of practice and understanding of framework is missed. The result of such missed opportunities may be that prospective teachers are not well-positioned to evaluate, design and deliver appropriate educational courses themselves. Given that an OBE approach is increasingly a part of primary and secondary education in Hong Kong, this is of special concern to the focus in this study.

A preliminary but viable premise may be drawn from these findings: in implementing an OBE initiative in a teacher education program, students have to be explicitly informed of OBE principles and their intended impact on planning and implementation. This must occur for at least two reasons, as a pre-requisite to their making informed evaluations regarding the quality of OBE innovations and as an enhancement to their capacity to become educational professionals.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Our preliminary finding was that the SEOBLS, with some enhancement may provide a viable framework for informing course enhancement from our students' perspective. However, a lack of transparency in providing students with an awareness of the OBE innovation may have had a negative effect on students' ability to fully judge the planning and changes that OBE represented in this course. This is of significance to teacher education in general, as all teacher education programs strive not just to build content experts, but those skilled in the practice of pedagogy. Transparency is important, as well in a more general sense for evaluating curricular change in teacher education and more broadly, tertiary education. As long as students' evaluation is deemed an important element of understanding the quality of change, we must assume some accountability for building students' capacity to make a fully informed evaluation. Designing greater transparency into an OBE process and more generally, any curricular change process would seem an important step in enhancing the value and quality of student evaluation.

Also important is the significance of conducting research into an educational initiative. Rather than moving straight from theory into adoption, HKIEd chose to engage in a pilot study. As a result, the Institute has been able to better chart a course forward and has taken initial but important steps in the application of methodologies for doing so; specifically the use and refinement of an survey focusing on both students' perceptions and the nature of our change initiative, the use of a positively packed scale to allow for sufficient discrimination with specific populations, and the analysis of data using both factor and Rasch analysis to gain a rich and valid picture of students' perceptions of an OBE-initiative. Too often, research is a missing link in an institution's movement from theory into implementation. The research steps taken in this study steps could be applied in a variety of settings to a variety of educational initiatives, providing informed judgements towards useful change.

References

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York: Routledge.

Baron, M. A., & Boschee, F. (1996). Dispelling the myths surrounding OBE. *Phi Delta Kappan*, *77*(8), 574-576.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Barr, R. B., & Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education." *Change, 27*(6), 13–25.

Biggs, J. & Tang, C. (2007). *Teaching for Quality Learning at University* (3rd ed.). Buckingham: SRHE and Open University Press.

Bond, M. H., & Hwang, K. K. (1986). The social psychology of Chinese people. In M. H. Bond (Ed.), *The psychology of Chinese people* (pp. 213-266). Oxford, UK: Oxford University Press.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: LEA.

Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports, 94*, 1015-1024.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1-38.

Deneen, C. (2009, December). *Global standards, local solutions: OBA and academic quality enhancement at HKIEd*. Paper presented at the 6th UGC Symposium on 3+3+4: Enhancing and Assessing Students' Learning Outcomes for the New 4-year Curriculum, Hong Kong, SAR.

Ewell, P. (2005). *Applying student learning outcomes to higher education: An overview*. Report prepared for the Hong Kong University Grants Committee.

Greimel-Fuhrmann, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality - analysis of relevant factors based on empirical evaluation research. *Assessment & Evaluation in Higher Education, 28*, 229-238.

Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for Social Sciences* (pp. 301-315). Thousand Oaks, CA: Sage.

Irving, S. E. (2004). *The development and validation of a student evaluation instrument to identify highly accomplished mathematics teachers.* Unpublished doctoral dissertation, University of Auckland, Auckland, NZ.

Joint Committee on Standards for Educational Evaluation (1988). *The Personal Evaluation Standards*. Newbury Park: Sage.

Jöreskog, K. G. (2007). Factor analysis and its extensions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 47-77). Mahwah, NJ: LEA.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*

Kember, D., & Wong, A. (2000). Implications for evaluation from a study of students' perceptions of good and poor teaching. *Higher Education, 40*, 69-97.

Killen, R. (2007). *Teaching Strategies for Outcomes-Based Education* (2nd ed.). Cape Town: Juta.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

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.

Linacre, J. M. (2009). WINSTEPS Rasch measurement (Version 3.68. 0)[Computer program]. Chicago, IL: Winsteps.com.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341.

Nusche, D. (2007). *Assessment of Learning Outcomes in Higher Education: A comparative review of selected practices*. OECD Education Working Paper Number 15. Paris: OECD.

Richardson, J.T.E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387-415.

Stone, M. (December, 2005). *Opening Remarks.* Symposium on Outcome-based Approach to Teaching, Learning and Assessment in Higher Education: International Perspectives. The Hong Kong Polytechnic University. Hong Kong, SAR

Spady, W. (1994). *Outcomes-based education: Critical issues and answers*. Arlington, VA: American Association of School Administrators.

Wang, L. (2009). *Syllabus for ENG 1244, Introduction to Language Studies* (available from the Hong Kong Institute of Education).

Wang, L. (2010). Designing and Implementing Outcome-Based Learning in a Linguistics Course: a Case Study in Hong Kong. Procedia - Social and Behavioural Sciences (p. 10). Kyrenia: Elsevier.

Watkins, D. A., Marsh, H. W., & Young, D. (1987). Evaluating tertiary teaching: A New Zealand perspective. *Teaching and Teacher Education, 3*, 41-53.

Webb, T. (2009, May). *OBA and the Quality of Student Learning*. Paper presented at UGC Symposium "The Outcome-based Approach: Where are we now and where are we going?" Hong Kong, SAR.

Wiggins, G., & McTighe, J. (2005). *Understanding by design*. (2nd ed). Alexandria, VA: Association for Supervision and Curriculum Development.

Current citation: Deneen, C. C., Brown, G. T. L., Shroff, R. H., & Bond, T. G. (2012, accepted). Telling the Difference: A First Evaluation of an Outcome-Based Learning Innovation in Teacher Education. *Asia Pacific Journal of Teacher Education.*