Title: They can't tell the difference: A first evaluation of an Outcome-Based Learning innovation in teacher education

Authors: Christopher C. Deneen, Hong Kong Institute of Education, Deneen@ied.edu.hk
Gavin Thomas Lumsden Brown, Hong Kong Institute of Education, gtlbrown@ied.edu.hk
Trevor G. Bond, James Cook University, tbond@ied.edu.hk
Ronnie H. Shroff, Hong Kong Institute of Education, rshroff@ied.edu.hk

Abstract: Outcomes Based Learning (OBL) approaches to specifying curriculum, assessment, and teaching in higher education are intended to clarify and improve higher education students' outcomes. OBL is a major initiative being implemented in universities, worldwide, with widespread backing by governments and standards bodies.

At the HK Institute of Education, a pilot of OBL in teacher education classrooms was studied by surveying students' perceptions of the efficacy of the OBL innovation adopted in one course (n=85).

A survey questionnaire, designed to obtain evaluation of the OBL class and to make comparisons to non-OBL classes, was administered at the end of a one-semester class. Two scales were identified with exploratory & confirmatory factor analysis and Andrich's rating scale model. Student responses to the class evaluation scale and the comparative scale were slightly positive, indicating students did not perceive OBL approaches negatively, nor did they perceive significant differences between OBL and other course approaches previously experienced. Interview of the course instructor revealed that there was little explicit discussion of OBL with students.

We conclude that students have to be explicitly informed of OBL principles as a pre-requisite to their making informed evaluations regarding OBL efficacy. Implications for OBL adoption will be discussed.

Key Words/Terms: Outcome-Based Learning, Higher Education, Quality Assurance, Student Evaluation, Teacher Education

Introduction

Quality assurance and accountability systems are a recognized part of higher education (Harvey & Newton, 2007). Although much research within quality assurance has focused on US and European institutions (Nusche, 2008), studies are arising in Pacific Rim universities, especially Hong Kong (Deneen, 2009; Webb, 2009). In Hong Kong, there have been rapid and decisive transitions in how quality is defined, arrived at, and measured in higher education. There is systemic emphasis on defining quality by focusing on the learner through the adoption of an outcome-based approach. Outcome-based approaches to student learning (OBASL) have enjoyed both popularity and criticism worldwide. For example, OBASL is accused of being an external imposition or reductionist approach (Hussey & Smith, 2003) with significant and negative consequences for learners at multiple levels (Donnelly, 2007). Furthermore, there are few peer-reviewed empirical studies that demonstrate the effectiveness of OBASL (Brady, 1996). Those that do exist show some promise, however for the merit of an OBASL approach (Ginns, Prosser, & Barrie, 2007; Prosser et. al., 2007). The Hong Kong University Grants Committee has adopted OBASL as a key to improving teaching and learning quality (Ewell, 2006). The Hong Kong Institute of Education (HKIEd) is experimenting with innovations in OBASL and this paper reports a pilot study to develop methods of inquiry and identify initial impact of an OBASL innovation.

Background

A Shift within Quality Assurance towards the Learner

Traditionally, quality assurance has focused on inputs and outputs, such as recruitment demographics, graduate placement, and acquired research funding (Nusche, 2008). While these are clearly important indicators, focusing solely on input and output fails to establish what happens to learners. The use of student evaluations of teaching (SETs) partly addresses this question (Gravestock, & Gregor-Greenleaf , 2008), but does not guarantee quality. Increasingly, higher education is looking to learner experiences and achievements to evaluate quality (Kember, 2009). What students can do, know, or understand by the culmination of their educational experience (i.e., their outcomes) along with their perceptions of teaching quality may provide a more robust basis for determining the quality of higher education.

OBASL goes by many names and acronyms; this paper uses OBASL as a broadly encompassing term, inclusive of outcomes based learning (OBL), outcomes based education (OBE), and so on. OBASL has been characterized as both learner-centered and results-oriented (Baron & Boschee, 1996). Spady (1994) defines OBASL as an inherently systematic approach, in that it both focuses and interconnects the components of an educational system around the anticipated achievement of learners. Consequently, OBASL may play a significant role in informing a systematic, learner-centric approach to quality enhancement and assurance (Webb, 2009).

Exhibitions of learner achievement within an OBASL context constitute a "culminating demonstration" (Spady and Marshall, 1991. pp. 67), which, at minimum possesses two characteristics. The first characteristic is that intended outcomes have been set and promoted which clearly articulate achievement in terms of content and skills. The second characteristic governs the context of the demonstration, itself, which must enable visible demonstration of the learner-centered outcomes. Meaning is achieved through achieving an alignment of the all elements of the curriculum (i.e., teaching and learning activities, readings, assignments, assessments, etc.) with intended learning outcomes (Biggs and Tang, 2006). These two criteria provide guidance in constructing an OBASL educational experience, such as a course. Specifically, an OBASL course must consist of the planning, teaching, learning, and assessment that focus on the intended outcomes and which support the visible demonstration of learner centered-achievement.

Methods

At HKIEd, OBASL has involved staff development combined with a pilot innovation conducted in the Fall semester of 2009 under the auspices of the OBL Unit. The purpose of the pilot was to assess students' and lecturers' perceptions of the efficacy of the OBL innovation adopted in one course. After conducting a literature review, the OBL Unit worked with teaching staff to design course intended outcomes, appropriate activities, and assessments. Further, a student survey instrument was designed and administered to all students in a 12 week undergraduate B.Ed. course restructured to OBASL standards.

Items for the student survey instrument were written to cover three key aspects of OBASL (i.e., course intended learning outcomes, teaching & learning activities, and assessment) as well as elicit an overall evaluation of the course by comparing the OBASL course with other courses the students had taken. Further, these items were also classified as to whether they focused on the propriety, feasibility, utility, and accuracy aspects of the OBASL innovation. These four domains are the main evaluation criteria against which the merit and worth of an object can be determined (JCSEE, 1998). Propriety has to do with proper conduct of procedures, understanding that rights and interests of all involved stakeholders have to be respected. Feasibility has to do with the degree to which procedures were realistic, viable, and practical, given constraints on time and resources. Utility has to do with usefulness of procedures for the intended users. Accuracy has to do with clarity and validity and reliability of the OBL procedures.

Participants were asked to indicate how strongly they agreed or disagreed with each statement using a six-point, positively-packed (Lam & Klockars, 1982) agreement scale. 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 this skewed response scale is useful when participants are likely to agree with all statements, because the greater range of options within the positive range elicits greater variation in responses than when only two response points are used to capture positive orientation.

Sample

The sample consisted of students (*n*=85) in a 12 week undergraduate B.Ed. course.

Analysis

Two techniques were used to analyze the student responses: Rasch analysis and factor analysis. Rasch analysis is ideal for determining whether items belong to a single dimension (Bond & Fox, 2007), while factor analysis, using both exploratory and confirmatory approaches, is ideal for determining the number of dimensions in a data set.

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 (Infit and outfit mean squares; $0.75 < x < 1.3$) or transformed in to probability statistics (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 the residual patterns for evidence of second (or further) dimensions. Inspection of the Category response curves is adopted to verify that respondents used the rating scale response options in a meaningful, hence measurable, way.

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 behavior. 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 on the data by constraining items to be explained only by their respective factor (Jöreskog, 2007). 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).

Results

Rasch Findings

Initial analyses using Winsteps (Linacre, 2009) identified the difficulty of endorsement and fit characteristics of the items. Table 1 shows the item statistics in ‘measure order’ with the most difficult items to endorse at the top. Three items (i.e., #14, 15 and 16) were the most difficult to endorse and had consistently poor misfit to the underlying single dimension associated with the remaining 14 items (i.e., items 1 – 13 + 17). Reverse scoring of these misfitting items did not alter the conclusion that these three items were not part of the same underlying measurement construct.

Table 1. Item Rasch fit statistics in measure order.

					Infit		Outfit		
Item	Raw Score	*N*	Measure order	*se* (logit)	Mean Square	*z*	Mean Square	*z*	r_{pb}
16	299	89	1.57	0.13	1.61	3.7	1.71	4.2	0.38
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
17	351	89	0.64	0.14	0.87	-0.9	0.89	-0.7	0.66
10	364	89	0.4	0.14	0.72	-2.1	0.72	-2.1	0.68
11	381	89	0.07	0.14	1.16	1.1	1.25	1.6	0.60
2	385	89	-0.01	0.14	0.75	-1.8	0.76	-1.7	0.75
13	385	89	-0.01	0.14	0.70	-2.2	0.71	-2.1	0.73
8	392	89	-0.15	0.14	0.53	-3.8	0.55	-3.6	0.8
12	399	89	-0.29	0.14	0.74	-1.9	0.77	-1.6	0.73

7	401	89	-0.33	0.14	0.78	-1.5	0.77	-1.6	0.76
4	400	88	-0.41	0.15	0.69	-2.3	0.7	-2.2	0.75
5	408	89	-0.48	0.15	0.88	-0.8	0.93	-0.4	0.72
1	410	89	-0.52	0.15	0.45	-4.6	0.48	-4.3	0.86
9	418	89	-0.7	0.15	0.90	-0.6	0.90	-0.7	0.77
3	419	89	-0.72	0.15	0.97	-0.1	0.97	-0.1	0.72
6	423	89	-0.81	0.15	1.02	0.2	1.03	0.2	0.76
M	383.1	88.9	0	0.14	0.98	-0.4	1.01	-0.3	
SD	33.3	0.2	0.65	0.01	0.44	2.8	0.45	2.8	

While 54% of variance was explained by the main scale (i.e., 14 items), almost one quarter (23.9%) of the unexplained variance could be attributed to those three misfitting items. Since such a large proportion of variance is accounted for, rather than random noise if only one dimension had been present, the conclusion is that at least two dimensions are present in the data.

Figure 1 shows the rating scale model analysis of the probability distributions relative to difficulty of endorsement for the six rating scale responses. The relatively uniform peaks for the four (non-extreme) categories suggests that all options exhibited clear increments in difficulty of endorsement across all respondents. The uniform shape of the response curves across the measurement scale provides evidence for the the appropriateness of the positively-packed response options as an interval level measurement scale.

Quite importantly, the items retained in this scale did not show spread across the entire measurement scale relative to the spread of the students on the same scale. Students, on average, found the items comparatively easy to endorse, with item difficulties clustered around the mid-point of the scale.

EFA Findings

One missing value was imputed using the expectation maximization procedure (Dempster, Laird, & Rubin, 1977). Exploratory factor analysis (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. 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 comparing workload, speed, and difficulty of the OBL course to other courses (i.e., items 14, 15, and 16). Confirmatory factor analysis had marginal fit to the 2-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 that had strong cross-loadings to other items or factors (i.e., #16, 1, and 11). By trimming these three items, the 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).

The inter-correlation between factors was weak (*r*=.11) indicating relative independence of the two factors. The mean score for Factor 1 was 4.45 (*SD*=.76), while the mean for Factor 2 was 3.80 (*SD*=1.01). The difference in means had a large Cohen's (1992) effect size (*d*=.72). This means students agreed reasonably moderately with the general course evaluation and considerably less with the difficulty factor. In other words, this course did everything the students expected and wasn't too hard.

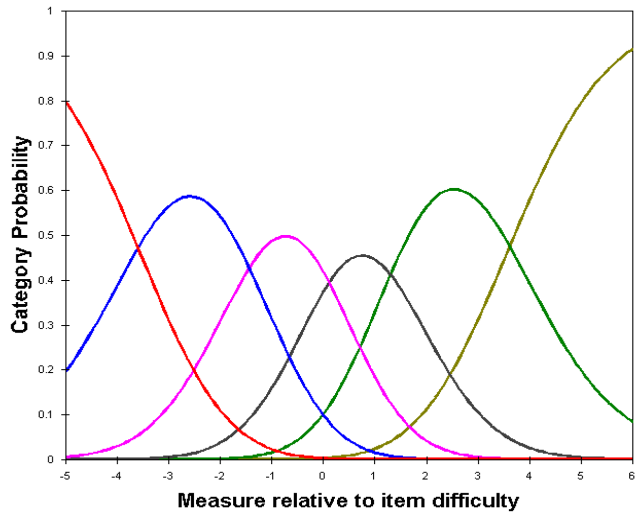

Figure 1. Rating scale model category probability curves for selected items.

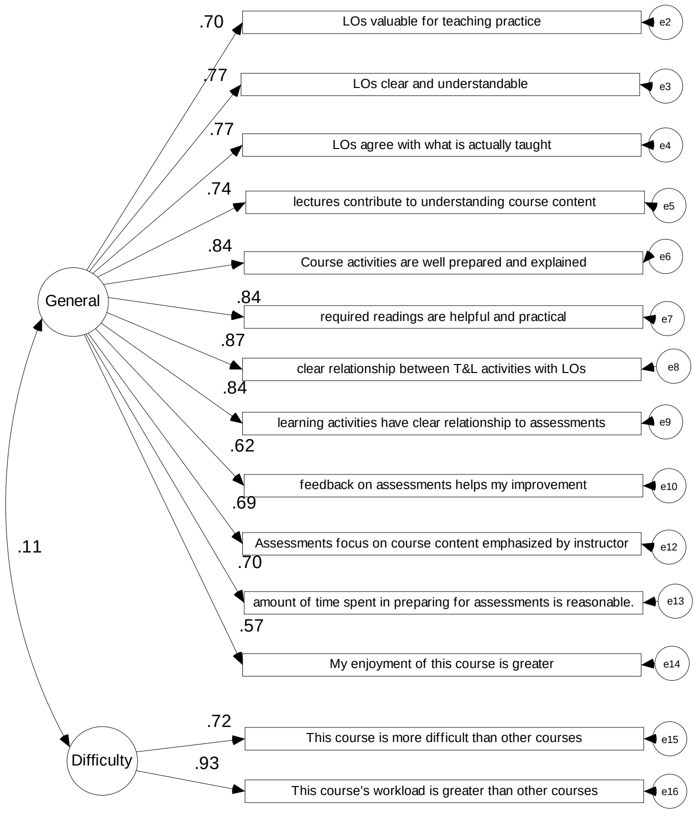

Figure 2. Two factor model of student responses to OBL course

Discussion

Refining mode of inquiry

One of the principal purposes of this pilot study was to refine the student survey inventory. Both sets of statistical analyses revealed that two scales were present in the data and that the positively-packed rating scale elicited equal interval responses from participants. The first factor, course evaluation, had many items, yet it was not possible to distinguish among the content areas or evaluation criteria, no doubt due to the low ratio of cases to items. The second factor, consisting of just two items which is generally considered insufficient, identified the relative difficulty of the course. This pattern indicates more items are needed for each intended category to increase the probability of detecting responses to each and more cases are needed to support more fine-grained analysis of responses to the student survey.

The use of both the Rasch and factor analytic approaches yielded very similar results. Both methods identified the same items as belonging to a second construct (i.e., items 14, 15, 16). Since the items were centered at the middle of the major scale, it would seem that simply averaging responses to those items may be a sufficient use of the measurement properties inherent in the rating scale and the items. Hence, this pilot has validated the response scale and pointed the way to new item development.

Substantive results

The two scales suggest that for these students their experience in the OBASL course was not a radical departure from a “regular” course. Students were, for the most part, satisfied with course quality in the areas we predicted as having maximum OBL impact, but neither did they consider the course easier or harder than other courses. Hence, it would appear students did not perceive the OBASL course as something different. There are several possible reasons for this result.

One possibility is that the actual intervention may have been quite limited. While considerable professional development support was given to the course tutors, there was no direct observation of the OBASL implementation. Student responses may have accurately reflected that in terms of teaching and learning activities and assessments there was no real difference in the OBASL course than any other course taught in the institution. Observational research, however, will require a great deal of delicacy and professional trust. Another possibility is that the instructors did not explain to the students that OBASL was actually being experimented with and hence, denied the students the opportunity to be aware of some innovation. If students are not aware of a possible difference, they may not note, report or reflect on such difference.

We conclude that students have to be explicitly informed that an OBASL approach is being attempted to make informed evaluations regarding its efficacy. This is especially the case because OBASL is founded on the notion of learner centered understanding of the educational process. Further, we conclude that fidelity of implementation is an essential characteristic that will require considerable tact in bringing about in a culture which treats academic freedom and independence as equivalent to never being evaluated.

This pilot moves us towards achieving a systematic means of enhancing and assuring learner-centered quality by developing an improved data collection tool and by pointing out the importance of student awareness and institutional transparency in making outcomes available to students.

References

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.

Biggs, J., & Tang, C. (2006). *Outcome-based teaching and learning- What is it? Why is it? How do we make it work?* A workshop document used in the City University of Hong Kong, Hong Kong Baptist University, and University of Hong Kong.

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

Brady, L. (1996). Outcome-based education: a critique. *The Curriculum Journal*, 7 (1), 5-16.

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 Polytechnic University, Hong Kong, SAR.

Donnelly, K. (2007). Australia's adoption of outcomes based education: A critique. *Issues in Educational Research*, 17(2), 183-206.

Ewell, P. (2006). *Applying student learning outcomes concepts and approaches at Hong Kong Higher Education Institutions: Current status and future directions*. National Center for Higher Education Management Systems.

Ginns, P., Prosser, M. & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: the perspective of currently enrolled students. *Studies in Higher Education*, *32*(5), 603-615.

Gravestock, P. & Gregor-Greenleaf, E. (2008). *Student Course Evaluations: Research, Models and Trends.* Toronto: Higher Education Quality Council of Ontario.

Harvey, L. & Newton, J. (2007). Transforming quality evaluation: moving on. In Westerheijden, D., Stensaker, B. & Rosa, M.J., (Eds.), *Quality Assurance In Higher Education: Trends in Regulation, Translation and Transformation* (pp. 225–245). Amsterdam: Springer.

Hussey, T. & Smith, P. (2003). The uses of learning outcomes. *Teaching in Higher Education,* 8 (3), 357-368.

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.

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.

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

Prosser, M., Martin, E., Trigwell, K., Ramsden, P. & Middleton, H. (2008). University academics' experience of research and its relationship to their experience of teaching. *Instructional Science,* 36, 3-16.

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

Spady, W. G. & Marshall, K. J. (1991). Beyond traditional outcome-based education. *Educational Leadership, 49* (2), 67-72.

Webb, T. (2009, May). *OBA and the Quality of Student Learning*. Paper presented at UCG symposium in Hong Kong. Hong Kong: Hong Kong, SAR.