CamEd Open Access Repository
ACBSP Region 10, 2023, pp. 95 - 104
DOI: https://doi.org/10.62458/CamEd/OAR/ACBSP/95-104

Outcome-Based Feedback: Collaborating with Students for Curriculum Review

Ly Monirith
CamEd Business School

The assessment of learning outcomes is vital in ensuring educational quality in the "fitness for purpose" model. The paper reports on the reliability and validity of the Outcome-Specific Questionnaire (OSQ) as a tool to improve learning outcomes. Survey data were collected from 1,210 undergraduate students and assessment data from three prominent institutions in Cambodia. Calibrated paired t-test results show significant increases of student learning from term start to end, and Cronbach's alpha results confirm high levels of reliability of all OSQs. There is strong evidence that the post-OSQ data for almost half of the courses correlate positively with the faculty assessments although most were conducted online during the COVID-19 pandemic. Together with qualitative data, it can be concluded that the OSQs tend to be valid as long as the assessments by faculty are valid and that their validity cannot be established when the assessments, as benchmarks, are not evidently valid. The study contributes to the minimal literature on and understanding of the rare practice of outcome-based evaluation and improvement of courses collaboratively by students and faculty in an outcome-based education system.

Keywords: *Learning outcomes, self-ratings, curriculum development*

INTRODUCTION

To provide context for the research study, outcome-based education will be introduced. Then, the rationale for the study will be provided. Next, two similar key terms will be defined. To complete the introduction to the study, the research objectives and questions will be laid out.

External quality assurance and accreditation agencies across the world are increasingly requiring outcome-based approaches to education (Kember & Ginns, 2012). This has occurred in a context of regional integrations and international mobility. The ASEAN University Network-Quality Assurance network was founded in 1998, and the first guidelines were published a few years afterwards (AUN-QA, 2016). The researcher was involved in developing the Quality Management of Educational Programs in Royal University of Phnom Penh and Royal University of Law and Economics manual published by the ASEAN University Network (Bin et al., 2016) to be handed to other universities in Cambodia aspiring to become members of the regional network. Cambodia's National Qualifications Framework was sub-decreed in 2014 by the Royal Government of Cambodia, and higher education institutions are working to ensure their curricula comply with the policy.

Spady (1994) is credited with coining the term "outcome-based education" (Willis & Kissane, 1997) and the associated movement (Glatthorn, 1993). Spady (1994) defined outcome-based education as

> … clearly focusing and organizing everything in an educational system around what is essential for all students to be able to do successfully at the end of their learning experiences. This means starting with a clear picture of what is important for students to be able to do, then organizing curriculum, instruction, and assessment to make sure this learning ultimately happens. (p. 1)

This definition comprises two main features: "backward design" of curriculum whereby expected learning outcomes are specified first before other curriculum components are developed (Wiggins & McTighe, 1998) and "constructive alignment" of curriculum contents, teaching and learning strategies, and student assessments to the specified outcomes (Biggs, 1999, 2014).

Rationale for the study

Evaluation or feedback is a key component of a continuous quality improvement cycle (Deming, 1982); however, in practice evaluation or feedback reports are often perfunctory, superficial, and ignored

(Newton, 2002). Thus, formal feedback systems need to be improved for curriculum development purposes. As part of the "fitness for purpose" quality model, an outcome-based curriculum must be evaluated to measure and improve its effectiveness in meeting its stated objectives. Yet, most course evaluation or feedback questionnaires are standard (generic) across courses, programs, and even institutions in the case of national surveys. Course evaluation usually seeks student feedback on inputs—resources, hours, facilities, textbooks, and other materials—and process—teaching and learning methods and activities. Banta (2008) remarks that there is barely any research on the linkage between process-based evaluation systems and the enhancement of learning outcomes. On the other hand, "Output measures increasingly dominate international quality frameworks as they speak to the dominant quality narratives of accountability …" (Marshall, p. 222). However, even so-called "outcome-based" questionnaires consist of generic competencies not specific to any discipline. As Schiekirka et al. (2014) found, "few evaluation tools directly assess learning outcomes for specific learning objectives" (p. 1).

Quality standards, such as those of the European Association for Quality Assurance in Higher Education, give significant weight to the engagement of stakeholders (Marshall, 2018). Stakeholders include faculty and students (Marshall; Reavill, 1998), but in university practice their inputs or feedback are usually not collected widely and systematically for curriculum review purposes. Typically, only a few instructor and student representatives are invited occasionally on an ad hoc basis. The formal process of collecting student feedback is often distilled from the day-to-day experiences of students and faculty (Harvey, 2011). The weights given to stakeholder groups need balancing, and "[t]he salience of the student as the definitive stakeholder needs to be genuinely valued" (Marshall, p. 345).

Definition of key terms

"Expected learning outcomes" are the knowledge, skills, and attitudes that the learners will be able to demonstrate (Anderson et al., 2005; McDonald, 1993). Each expected learning outcome must begin with an active verb to indicate an observable and assessable activity, not general verbs such as learn, have, be, know, and understand (Spady, 1994).

Program, course, teaching, or learning "objectives" are the knowledge, skills, and attitudes that the program, course, or teacher intends to provide, build, and cultivate. This paper uses the term objectives to refer to expected learning outcomes as stated on the course syllabus, while outcomes refers to the end-of-term achievements as perceived by the students themselves or assessed by the faculty.

Research objective and questions

This study assesses the reliability and validity of students' self-measures of learning outcomes in relation to course expected learning outcomes (CLOs):

1. Do students' perceptions of their beginning abilities against CLOs change after exposure to the course? If so, do their self-ratings increase or decrease? (reliability)
2. Do students significantly improve, during the course period, their self-reported outcomes against CLOs? If so, to what extent do they improve? (reliability)
3. Are students' post-test self-ratings correlated to the assessment by faculty? If so, how correlated are they? (validity)

LITERATURE REVIEW

The review of empirical literature will relate outcome-based education and higher education quality to the fitness for purpose model of quality and discuss the purposes of collecting student feedback.

Outcome-Based Education as fitness for purpose

Outcome-based education can be associated with a utilitarian purpose of education, which focuses on the usefulness or application of learning in the economy (Cheng, 2016) and relevance to professional practice (Harden et al., 1999) and everyday lives (Spady, 1994). Educational quality under this approach "refers to the degree of utility or impact" (Marshall, 2016, p. 215). Culminating outcomes must be purposeful (Willis & Kissane, 1997) and reflect adult life roles beyond schooling (Spady, 1994). Thus, outcome-based education can be linked to the fitness for purpose model of educational quality. The model is compatible with the backward design and constructive alignment features of outcome-based education—"the alignment of courses and programs to specific learning objectives and graduate attributes" (Marshall, 2018, p. 331). With the rising need for institutional accountability of resource allocations, outcome-based higher education is justified by assessing its quality against its utilitarian purpose.

Quality of higher education as fitness for purpose

Although definitions of quality in higher education vary (Harvey & Green, 1993; Marshall, 2016, 2018), the fitness for purpose model is "the most inclusive and least confronting" (Marshall, 2018, p. 331) and the most internationally upheld approach (Woodhouse, 1996). The model has spread to Southeast Asia, as evidenced by the adoption of outcome-based education by the ASEAN University Network Quality Assurance Network's Guide to AUN-QA Assessment at Program Level (Bin, 2015).

The fitness for purpose approach to quality assurance in higher education is rooted in business academia, including "management by objectives" (Cheng, 2016, p. 2). Higher education quality as fitness for purpose means that each institution must fulfil its own established purpose (Woodhouse, 1996) and serve the needs of all stakeholders (Cheng, 2016; Reavill, 1998). As stakeholder needs change over time, the purpose of education also changes, allowing for curriculum revisions.

Purposes of collecting student feedback

The purpose of student feedback has evolved from administrative utility to teaching and learning improvement (Leckey & Neill, 2001; Nair & Mertova, 2011). The former focuses on quality assurance or control of teaching performance, whereas the latter aims at the continuous enhancement of teaching and learning quality (Kember & Ginns, 2012; Harvey, 2011; Nair & Mertova, 2011) and the promotion of reflective practice for professional development among faculty (Leckey & Neill, 2001). Improving teaching and learning includes increasing the likelihood of students achieving the expected learning outcomes, and another purpose of student feedback is to support curriculum reviews (Kember & Ginns, 2012). In brief, evaluation or feedback is vital in closing the loop of any quality assurance or improvement cycle.

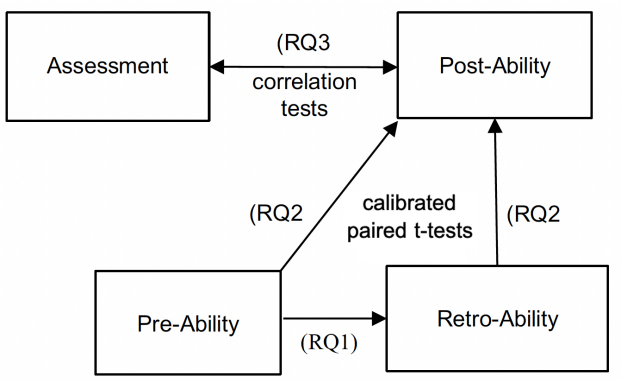
METHODOLOGY

Research design

The study utilizes calibrated paired t-tests and correlation tests. The research design and variables are displayed in Figure 1 and elaborated below.

Figure 1

Research Design and Variables



In each course, student abilities were derived as part of the calibration process to be explained in the Data Analysis section. Paired t-tests were conducted between Pre-Abilities and Retro-Abilities to determine whether students had changed their perceived beginning abilities after course exposure and to answer research question 1. Paired t-tests were conducted between Pre-Abilities and Retro-abilities against Post-Abilities, respectively, to identify whether there was significant learning over the term. The results help answer research question 2. To answer research question 3, students' post-ratings or Post-Abilities were correlation tested against the faculty assessment of those students' work to determine the content validity of the OSQ.

Content validity is indicated by correlation with other measures with "known or assumed" validity (Biggs & Telfer, 1987, p. 470). In this study, faculty assessments were assumed to be valid benchmarks for measuring the content validity of students' post-test self-ratings and, therefore, the Outcome-Specific Questionnaire (OSQ) instruments. For the correlation tests, the researcher developed Outcome-Specific Assessment Tools (OSATs) for the faculty members to align each part of their assessment tasks with the CLOs.

Data collection

Survey data were collected—mostly online during the COVID-19 pandemic—from 1,210 undergraduate students of eight courses at three prominent institutions in Cambodia. Assessment data were obtained from seven of the courses. The purposive sampling criteria include large enrolment per course for statistical purposes and assessable course-specific learning outcomes, i.e. knowledge and cognitive skills as classified by Cambodia's National Qualifications Framework. The reason for sampling

100- and 200-level courses with at least 100 students was to aim for sample size sufficiency per course. Furthermore, courses with objectives identified as rateable and assessable were sampled so that students would be able to rate their own abilities against the objectives and assessment tasks could be aligned to the objectives for correlation tests.

Instruments

The researcher developed the Outcome-Specific Questionnaires (OSQs) based on the expected learning outcomes of each course and were administered at the beginning (Pre-OSQ) and by the end of the term (Post-OSQ). Each OSQ began with informed consent and ended with demographic items. In the main part, the items were course-specific knowledge and cognitive skills from each course syllabus. The Pre-OSQ asked students to rate their own abilities at the time of survey, whereas the Post-OSQ asked them to rate their abilities retrospectively at term start (Retro items) and at the time of survey (Post items). The same rating scale was used in all OSQs: weak, moderate, good, and strong. An extract of a Post-OSQ is displayed in Table 1.

Table 1

Extract of a post-OSQ (Introduction to Environmental Science Course)

Please honestly rate your abilities at the semester start and now.	weak	moderate	good	strong
1a. describe rapid human growth as a fundamental environmental issue. (at term start)	○	○	○	○
1b. (now)	○	○	○	○
2a. identify sustainability concepts & its importance in conservation of resources. (at term start)	○	○	○	○
2b. (now)	○	○	○	○
...	○	○	○	○

Data analysis

Software applications utilized in data analysis include Excel, SPSS, and RStudio. The researcher completed the following data analysis stages per course:

1. Coding
2. Principal Component Analysis (PCAs) to determine items for each Principal Component to meet the unidimensionality assumption for calibration
3. Scale reliability (Cronbach's alpha) analyzes
4. Calibration of the items of each Principal Component with the Graded Response Model (GRM)
5. Calibrated paired t-tests to answer research questions 1 and 2
6. Pearson or Spearman correlation tests to answer research question 3

RESULTS

PCA results

Only accessible course-specific objectives were selected for the Outcome-Specific Questionnaires (OSQs). Principal Component Analysis (PCA) results confirm that all OSQ items of each sampled course belong to one Principal Component and, therefore, can be calibrated. Particularly, the first component was found to be the only one whose Eigenvalue is higher than 1.0 and the inflection point in the scree plot. The following figure and table display some PCA results of a sampled course (Khmer & Regional History).

Figure 2

Sample scree plot of pre-learning outcome components

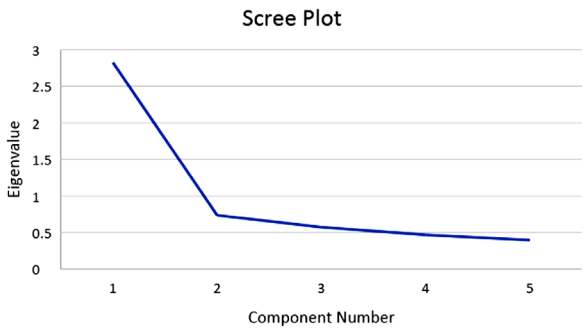


Table 2

Sample component matrix of pre-learning outcomes

Variable	Component 1
Pre3	0.78
Pre5	0.78
Pre2	0.77
Pre4	0.77
Pre1	0.66

Scale reliability results

Cronbach's alpha results below range from .72 to .96, which indicate high levels of scale reliability of the Outcome-Specific Questionnaires.

Table 3

Cronbach's Alpha results

Course	Items	Pre	Retro	Post
Khmer & Regional History	5	.81	.89	.79
English Reading & Composition I	18	.93	.95	.93
English Reading & Composition II	11	.89	.88	.89
Introduction to Environmental Science	10	.88	.93	.90
Cultural Anthropology	6	.90	.93	.91
Health Education & Fitness	4	.86	.83	.72
Introduction to Computers	11	.94	.96	.95
English for Academic Purposes 1	10	N/A	.92	.88

Paired T-Test results

Calibrated Pre- vs. Retro-Abilities

The calibrated paired t-test results indicate that in a simple majority of sampled courses there are non-significant differences between pre- and retro-perceptions. In two courses, there are significant decreases from pre- to retro-ratings with medium effect sizes. In one course, there is a significant increase from pre- to retro-ratings with a small effect size.

Calibrated Pre-/Retro- vs. Post-Abilities

The results show that in all sampled courses there are significant increases from retro- to post-abilities with three large, four medium, and one small effect sizes. Similarly, in almost all of the courses there are significant increases from pre- to post-abilities with one large, three medium, and two small effect sizes. Only in one course is there a non-significant increase from pre- to post-abilities. Due to space limitations, only typical results of the calibrated paired t-tests are shown in Table 4 (Introduction to Computers course).

Table 4

Sample Paired T-Test results

Statistic	Pre-Ability Z	Retro-Ability Z	Pre-Ability Z	Post-Ability Z	Retro-Ability Z	Post-Ability Z
Mean	-0.08	-0.15	-0.08	0.53	-0.09	0.55
Variance	0.98	1.22	0.98	0.81	1.23	0.79
Observations	107	107	107	107	140	140
Pearson Correlation	0.48		0.4		0.53	
t Stat	0.69		6.05		7.78	
P one-tail	0.25		0	sig. inc.	0	sig. inc.
P two-tail	0.49	non-sig. dif.	0		0	
Cohen's d effect size	0.07	small	0.58	medium	0.66	medium

Correlation results

The third research question asks whether students' post-test self-ratings are correlated to the assessment by faculty. Correlation tests result in either non-significant or inverse correlations in four of the seven courses whose assessment data were provided. In three courses, in contrast, the tests find positive correlations between Post-OSQ ratings and academic assessments. In the English Reading & Composition II course, there is a correlation with a Pearson r coefficient of 0.41 at the 5% significance level. The Pearson correlation test of the English for Academic Purposes 1 course results in a coefficient r of 0.27 at the 1% significance level. A Spearman correlation test had to be used in the Introduction to Environmental Science, and the result is a rho coefficient of 0.22 at the 5% significance level.

DISCUSSION

The calibrated paired t-test results add evidence of the reliability of the OSQs and indicate that the Post-OSQs (including retro- and post-items) can be used without the Pre-OSQs. Likewise, Schiekirka et al. (2014) compared and found no significant difference between pre-post gains and those from then-test (retrospective-test) to post-test.

In terms of the content validity of the OSQs, Post-OSQ self-reports and academic assessments are positively correlated at the 95% confidence level in two courses and 99% confidence in another course. Although over half of the courses see non-significant correlations, many lecturers expressed concerns that their assessments were unlikely valid because they were teaching and assessing students online for the first time. For instance, they were afraid that students could cheat during online tests and exams. The researcher had to remind the lecturers many times before they finally provided the assessment data. In the worst cases, one lecturer did not give any assessment data, and another lecturer withdrew from the study (the withdrawn lecturer's course

and his students are not reported in the course and participant counts).

It is also worth noting that most of the faculty members developed their own assessment tasks even if they taught the same courses. As outcome-based education was a recent reform in Cambodia, lecturers' experience and skills in developing course objectives and constructively aligning assessments with the objectives were limited; thus, their individually designed assessments might not have been as valid as assumed to be. As Fields (2019) points out, the lack of training in the science of assessment affects the reliability and validity of teacher-made tests.

The faculty members whose assessments were correlated with their students' self-ratings were the exceptions. One of the lecturers of the English Reading & Composition II course gave me access to his Google Forms exams, in which items had been well developed with objective item scoring. Conscious of the course's expected learning outcomes (CLOs), he aligned the items of his exams to specific CLOs. He was a member of an English teachers organization's leadership team responsible for placing international teachers in developing countries. The organization provides new and experienced teachers with ongoing professional development. The Introduction to Environmental Science professor sent me her exam paper and was able to align the exam items to specific CLOs. Her exam scores of 112 students were differentiated from 12 to 29 (out of 30) points, which I ranked into 16 categories of ordinal data paired with 81 students' calibrated self-rating data. None of the participating lecturers have access to students' self-rating data, so it is impossible for them to try to match the assessment and self-rating data in any way.

In the course with correlation at the 1% significance level, Cambodian and international faculty members held team meetings in which they actively discussed exam contents and rubrics to be used across all classes of the English for Academic Purposes 1 course. In the debates, they even referred to the course-specific objectives, e.g. when deciding whether to include sections other than essay writing in the exams. The internal quality assurance division of the school required that each exam section show alignment to specific CLOs. It is not surprising that this institution has earned recognition from many international quality assurance agencies.

Constructively aligning assessment tasks to CLOs means proactively matching them while designing the assessment tasks. This will increase the content validity of the assessments (Davidson & Lynch, 2002). Conversely, aligning assessment tasks to CLOs after the tasks are set will not improve the assessment validity or lack thereof. Despite being aware that the Pre- and Post-OSQs focus on CLOs, most participating faculty members did not offer evidence of their assessment tasks being already aligned to CLOs, for example on their course syllabi as required by the Department of Higher Education of the Ministry of Education, Youth and Sport. Many of them completed the Outcome-Specific Assessment Tools (OSATs) for the study only after many reminders, and some of them did not return the OSATs at all. In brief, the validity of most assessment tasks and scores provided cannot be established.

Chen and Foung (2019) experimented with structural equation modeling to evaluate lecturer-made assessments in terms of their alignment to course objectives. Contrasting a model created by exploratory factor analysis and hypothesized models created with input from course coordinators, they found the latter to directly and better illustrate the assessment-objective alignment. Similarly, in the current study, all participating lecturers were consulted for assessment-objective alignment with the OSAT forms. However, many of them offered qualitative judgements on the lack of validity of their online assessments.

Students must clearly understand the learning outcomes expected of them in order to give valid self-ratings. Even teachers themselves need to understand educational objectives so that "they will judge their students' learning more validly and reliably" (Willis & Kissane, 1997, p. 6). This lack of understanding of course objectives may explain the lack of correlations between self-reports and academic assessments in many courses.

Students' lack of experience assessing themselves might also account for the discrepancies. Students need practice in self-assessment as a grading mechanism (Tait-McCutcheon & Knewstubb, 2018) and in general, i.e. not linked to course grades. They need to develop skills in assessing their own abilities (Boud & Falchikov, 2007), as well as their progress (Cassidy, 2007). Boud and Soler (2016) argue that these methods of self-judgements are some practical forms of sustainable assessment habits for learning purposes beyond the course.

Although overall grades are insufficient in assessing and documenting actual learning outcomes (Anderson et al., 2005; Mabin & Marshall, 2012), this

research study does not rely on overall course scores. On the contrary, assessment tasks were selected based on individual student (not team) assessment type and alignment to course-specific objectives. In other words, data from only those qualified assessment tasks were used in correlation tests.

Some literature does not support a strong correlation between students' self-assessment and the lecturer's assessment of the students. According to Tait-McCutcheon and Knewstubb's (2018) study, for example, about 25% of students assessed their own work differently from their lecturer's assessment. In the larger population, however, students' grade expectations are generally close to the actual grades they get from the teachers (Marsh & Roche, 1997). Moreover, self-assessment is different from self-report. In self-assessment, students know their self-scoring may affect their overall grades. In the current study, in contrast, the questionnaires stated that self-ratings would not affect their grades and asked the participants to honestly rate their abilities.

Harvey (2011) suggests that feedback collection be customized to each course by the course team, that student feedback focuses more on the learning than teaching, and that doing these will encourage faculty to be less centered on their inputs and perspectives and more considerate of student activities and learning. A few researchers have acted in parallel to these recommendations. Measuring the correlation between student self-reports on their outcome-based questionnaire and those on a traditional course evaluation questionnaire, Raupach et al. (2012) conclude that an outcome-based instrument is "less heavily confounded by construct-irrelevant factors" (p. 8) than the generic—input- or process-based—student feedback tool. A similar statement can be made about the OSQs used in the current study: Principal Component Analysis (PCA) results confirm that all OSQ items of each sampled course belong to one Principal Component or construct, i.e. course-specific outcomes in the cognitive domain of Bloom's Taxonomy.

However, the OSQs are unique as they apply a rating scaling from weak to strong rather than a 7-point scale as used by Combs et al. (2008), the 6-point scale from fully agree to completely disagree as used by Raupach et al. (2012), or a 5-point Likert scale as used by Kaliannan and Chandran (2012). The reason is cultural context: Cambodians rarely express strong disagreement words such as completely disagree or strongly disagree so they can show a more balanced mentality—called upekkha (equanimity) in a Buddhist ethic of care.

Although Kaliannan and Chandran (2012) conducted pre- and post-surveys based on course outcomes, they reported only descriptive statistics without discussing the validity of their instrument. Raupach et al. (2012) and Schiekirka et al. (2014) used mean ratings in calculating comparative self-assessment gains from pre- to post-tests. Schiekirka et al. (2014) summarize, "Estimating learning outcome from comparative student self-ratings is a reliable and valid method" to determine strengths and weaknesses in undergraduate courses (p. 1). However, both research teams treated ordinal data as numerical data without any calibration. The current study contributes to the literature on the reliability and validity of outcome-based questionnaires by applying calibration of rating data after running PCAs to assure fit of the Graded Response Model.

CONCLUSION

The study contributes to course development and quality assurance. The evaluation of a course rarely focuses on the outcomes. When learner feedback is utilized to improve course objectives, in an outcome-based curriculum all other course components must be aligned accordingly to improve learning and, therefore, better meet the needs of students, faculty, and the program. In brief, valid feedback is necessary in the evaluation of a course as it closes the gap of course-level curriculum development.

In terms of internal quality assurance, a feedback instrument facilitates sustainable, systematic inputs for outcome-based curriculum reviews in continuous quality improvement cycles. Evaluation is a key component of a continuous quality improvement cycle (Deming, 1982). This study demonstrates how an instrument may improve formal feedback systems. Such improvement also provides documentation for external quality assurance purposes such as accreditation and accountability.

RECOMMENDATIONS

All internal and external stakeholder groups should be involved in the outcome-development process, which will enable them to have shared understanding of and commitment to the educational objectives (Darling-Hammond, 1993). Faculty must ensure that their students understand the expected outcomes from their first session. Program administrators and training providers could customize the Outcome-

Specific Questionnaires to various courses and workshops by substituting the expected learning outcomes and, if they prefer, the rating scale. Survey data analysts must calibrate rating data, and the Graded Response Model fits this psychometric purpose.

REFERENCES

Anderson, H. M., Moore, D. L., Anaya, G., & Bird, E. (2005). Student learning outcomes assessment: A component of program assessment. *American Journal of Pharmaceutical Education, 69*(2), 256-268.

AUN-QA. (2016). *AUN-Quality Assurance timeline.* http://aun-qa.org/briefintroduction

Banta, T. (2008, November). *Profiles of good practice in assessing student learning outcomes.* Keynote address. ATE and HELT Conference, Victoria University of Wellington.

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

Biggs, J. (1999). *Teaching for quality learning at university.* Open University Press.

Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education, 1*, 5-22.

Biggs, J. B., & Telfer, R. (1987). *The process of learning* (2nd ed.). Prentice-Hall of Australia.

Bin, J. O. C. (2015). *Guide to AUN-QA assessment at program level* (version 3.0). ASEAN University Network.

Bin, J. O. C. et al. (2016). *Quality management of educational programs in Royal University of Phnom Penh and Royal University of Law and Economics.* ASEAN University Network.

Boud, D., & Falchikov, N. (2007). Developing assessment for informing judgment. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment for Higher Education: Learning for the Longer Term* (pp. 181-197). Routledge.

Boud, S., & Soler, S. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400-413. https://doi.org/10.1080/02602938.2015.1018133

British Accreditation Council. (2017). *Competitive collaboration – An overview of the INQAAHE conference: "Between collaboration and competition: The promises and challenges for quality assurance in higher education."* http://www.the-bac.org/2017/03/06/competitive-collaboration-an-overview-of-the-inqaahe-conference-between-collaboration-and-competition-the-promises-and-challenges-for-quality-assurance-in-higher-education/

Cassidy, S. (2007). Assessing "inexperienced" students' ability to self-assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education, 32*(3), 313-330.

Chen, J., & Foung, D. (2019). Connecting teacher-made assessment through learning analytics: An empirical model. In E. White & T. Delaney (Eds.), *Handbook of Assessment Literacy and Teacher-Made Assessment in the Language Classroom*. https://www.igi-global.com/gateway/chapter/217148

Cheng, M. (2016). *Quality in higher education: Developing a virtue of professional practice. Sense.*

Coates, H. (2011). Tools for effective student feedback. In C. S. Nair & P. Mertova (Eds.). *Student feedback: The cornerstone to an effective quality assurance system in higher education* (pp. 101-118). Chandos.

Combs, K. L., Gibson, S. K., Hays, J. M., Saly, J., & Wendt, J. T. (2008). Enhancing curriculum and delivery: Linking assessment to learning objectives. *Assessment & Evaluation in Higher Education, 33*(1), 87-102.

Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan, 74*(10).

Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications.* Yale University Press.

Deming, W. E. (1982). *Out of the crisis.* Cambridge, MA: MIT Center for Advanced Engineering Study.

European Commission, & New Zealand Qualifications Authority. (2016). *Comparative analysis of the European Qualifications Framework and the New Zealand Qualifications Framework: Joint technical report.* https://www.nzqa.govt.nz/assets/About-us/Our-role/international-ed/EQF-NZQF-compa

rative-analysis/3254-NZQA-Comparability-EC-NZQA-Joint-Report-FINAL-R2-online-version.pdf

Fields, M. (2019). Common errors in teacher-made test design. In E. White & T. Delaney (Eds.), *Handbook of Assessment Literacy and Teacher-Made Assessment in the Language Classroom* (pp. 328-346). https://www.igi-global.com/gateway/chapter/217160

Glatthorn, A. A. (1993). Outcome-based education: Reform and the curriculum process. *Journal of Curriculum and Supervision, 8*(4), 354-363.

Harden, R. M., Crosby, J. R., & Davis, M. H. (1999). AMEE guide no. 14: Outcome-based education: Part 1—An introduction to outcome-based education. *Medical Teacher, 21*(1), 7-14.

Harvey, L. (2011). The nexus of feedback and improvement. In C. S. Nair & P. Mertova (Eds.). *Student feedback: The cornerstone to an effective quality assurance system in higher education* (pp. 3-26). Chandos.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education, 18*(1), 9–34. https://doi.org/10.1080/0260293930180102

Houston, D. (2007). TQM and higher education: A critical systems perspective on fitness for purpose. *Quality in Higher Education, 13*(1), 3-17. https://doi.org/10.1080/13538320701272672

Kaliannan, M., & Chandran, S. D. (2012). Empowering students through outcome-based education (OBE). *Research in Education*, (87), 50-63.

Kember, D., & Ginns, P. (2012). Evaluating teaching and learning: *A practical handbook for colleges, universities and the scholarship of teaching. Routledge.*

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice, 41*(4), 212-218.

Leckey, J., & Neill, N. (2001). Quantifying quality: The importance of student feedback. *Quality in Higher Education, 7*(1), 19-32. https://doi.org/10.1080/13538320120045058

Mabin, V. J., & Marshall, S. J. (2012). Beyond assessment: Assuring student learning in higher *education. Proceedings of the Symposium on Assessment and Learner Outcomes,* 187-203.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187-1197.

Marshall, S. J. (2016). Quality as sense-making. *Quality in Higher Education, 22*(3), 213-227. https://doi.org/10.1080/13538322.2016.1263924

Marshall, S. J. (2018). Shaping the university of the future: *Using technology to catalyse change in university learning and teaching. Springer Nature Singapore*.

McDonald, J. P. (1993). Three pictures of an exhibition: Warm, cool, and hard. *Phi Delta Kappan, 74*(6), 480-485.

Nair, C. S., & Mertova, P. (Eds.). (2011). Student feedback: *The cornerstone to an effective quality assurance system in higher education.* Chandos.

Newton, J. (2002). Views from below: Academics coping with quality. *Quality in Higher Education, 8*(1), 39–61.

Raupach, T., Schiekirka, S., Munscher, C., Beissbarth, T., Himmel, W., Burckhardt, G., & Pukrop, T. (2012). Piloting an outcome-based program evaluation tool in undergraduate medical education. *GMS Z Med Ausbild, 29*(3), Doc44. https://doi.org/10.3205/zma000814

Reavill, L. R. P. (1998). Quality assessment, total quality management and the stakeholders in the UK higher education system. Managing Service Quality: *An International Journal, 8*(1), 55-63. https://doi.org/10.1108/09604529810199395

Royal Government of Cambodia. (2014). *Sub-decree on Cambodian national qualifications framework. Author.*

Saunders, D. B. (2011). *Students as customers: The influence of neoliberal ideology and free market logic on entering first-year college students* (Open access dissertations. Paper 377). http://scholarworks.umass.edu/open_access_dissertations/377

Schiekirka, S., Anders, S., & Raupach, T. (2014). Assessment of two different types of bias affecting the results of outcome-based evaluation in undergraduate medical education. *BMC Medical Education, 14*(149), 1-9. https://doi.org/10.1186/1472-6920-14-149

Schwarz, G., & Cavener, L. A. (1994). Outcome-based education and curriculum change: Advocacy, practice, and critique. *Journal of Curriculum and Supervision, 9*(4), 326-338.

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

Spronken-Smith, R., Bond, C., McLean, A., Frielick, S., Smith, N., Jenkins, M., & Marshall, S. (2015). Evaluating engagement with graduate outcomes across higher education institutions in Aotearoa/ New Zealand. *Higher Education Research & Development, 34*(5), 1014-1030. https://doi.org/10.1080/07294360.2015.1011098

Tait-McCutcheon, S., & Knewstubb, B. (2018). Evaluating the alignment of self, peer and lecturer assessment in an Aotearoa New Zealand pre-service teacher education course. *Assessment & Evaluation in Higher Education, 43*(5), 772-785. https://doi.org/10.1080/02602938.2017.1408771

Wiggins, G., & McTighe, J. (1998). *Understanding by design. Association for Supervision and Curriculum Development.*

Willis, S., & Kissane, B. (1997). *Achieving outcome-based education: Premises, principles and implications for curriculum and assessment.* Australian Curriculum Studies Association.

Woodhouse, D. (1996). Quality assurance: International trends, pre-occupations and features. *Assessment and Evaluation in Higher Education, 21*(4), 347–356. https://doi.org/10.1080/0260293960210405