CONVERGENT VALIDITY OF INVIGILATED, SUMMATIVE ONLINE ASSESSMENTS IN BUSINESS MATHEMATICS COURSES AND STUDENTS' COMPARATIVE PERFORMANCE IN ONLINE AND OFFLINE ASSESSMENTS

Victor K. Y. Chan and Hio Man Lei
Macao Polytechnic University
Rua de Luís Gonzaga Gomes, Macao SAR, China

ABSTRACT

The validity and specifically the convergent validity of a course's online assessments (i.e. the extent to which the online assessments really measure what is measured by other assessments meant to measure the constructs of the course's learning outcomes) so far receive little attention, especially for business mathematics courses. Based on a business mathematics course at the authors' university, this article aims to verify the convergent validity of an invigilated, summative online assessment by evaluating the extent to which the online assessment measures what is measured by an invigilated, summative offline assessment meant to measure the construct of the course's learning outcomes. In addition, this article attempts to deduce whether the students performed better in the online assessment than in the offline assessment or vice versa. Findings was that such convergent validity was just scarcely acceptable and that students in the online assessment drastically outperformed themselves in the offline assessment. Reasons are proposed for the findings, for example, students' computer anxiety, perceptions of and attitudes towards online assessments as well as some distinctive features of online assessments in mathematical courses.

KEYWORDS

Summative Online Assessment, Invigilated, Comparative Performance, Mathematics Courses, Offline Assessment.

1. INTRODUCTION

Alongside the phenomenal proliferation of e-learning applications in universities and other tertiary education institutions across the globe over the past two decades, summative online/electronic assessment of students (e.g. through tests and examinations on electronic e-learning platforms) has been piloted, launched and even institutionalized in the realm of tertiary education. Vis-à-vis offline/traditional (pen-and-paper) counterparts, online assessments are inherently superior in the sense of:

- cost and time savings in view of the automated administration, grading/marking/scoring and storage of students' works,
- possibly customised and/or immediate feedback to students, which is of pedagogical benefit,
- enhanced student engagement due to the novelty and appeal of this assessment modality,
- geographical flexibility of students in submitting their works on campus or off campus and
- minimised human errors in grading/marking/scoring.

(Hewson, Charlton and Brosnan, 2007; Hewson, 2012).

Nevertheless, ever since their advent, the validity of online assessments has loomed large and been controversial among tertiary educationists (Hewson, 2012). Validity refers to the extent to which the online assessments in a particular course (or module dependent on the specific terminology adopted in a particular tertiary institution) turn out to measure what they are meant to measure, i.e. specifically the learning outcomes of the course (Dennick, Wilkinson and Purcell, 2009; Whitelock, 2009; Hewson, 2012). Traditionally and generally, validity of any instrument to measure any construct (i.e. abstract variable), be the construct a course's learning outcomes or otherwise, covers hierarchical perspectives as follows:

- content validity
    - face validity
- criterion-related validity
    - concurrent validity
    - predictive validity
- construct validity
    - convergent validity
    - discriminant/divergent validity

(Heale and Twycross, 2015).

In particular, convergent validity concerns the extent to which the instrument really measures what is measured by other instruments meant to measure the construct (Heale and Twycross, 2015). Therefore, for an online assessment (as an instrument to measure the construct of a course's learning outcomes) to achieve high convergent validity, it should at least measure what is measured by an offline assessment broadly recognised as an instrument to measure the construct of the course's learning outcomes. In other words, such high convergent validity can be manifested by consistency between the online assessment scores and the offline assessment scores when both assessments are administrated to the same student sample.

To the knowledge of the authors, consistency between online and offline assessment scores in the form of a high correlation coefficient or otherwise for the purpose of substantiating convergent validity of an online assessment has still not been conclusively established generally across multiple disciplines. This is especially true of the business mathematics discipline. As a matter of fact, there are appreciable empirical studies hitherto on validity of online assessments and effectiveness of e-learning specific to courses in such disciplines as psychology (Hewson, Charlton and Brosnan, 2007; Hewson, 2012) and medicine (Pei and Wu, 2019; Hope *et al.*, 2021). In contrast, comparable studies in business mathematics disciplines are relatively the minority. As Pei and Wu (2019) contend, curriculum types (and thus disciplines) may dictate the outcomes of e-learning (and thus online assessments), so it is pointless to miss out business mathematics disciplines as such. In fact, online assessments in mathematical disciplines may deserve particular heed in the sense that not only are online assessments most prevalently adopted in mathematical (or numerate) disciplines (Hewson, 2012) but also they are characterised by the following distinctive features setting them apart from other disciplines in the context of online assessments:

(a) On top of serving solely as content repositories, content retrieval systems and operational media for the assessment and learning of any disciplines, computer platforms for online assessment and e-learning at large are direct assistive tools for the computation-intensive content of mathematical disciplines (Hussain *et al.*, 2014).

(b) Students may take advantage of such computer platforms to countercheck their answers during online assessments of mathematical courses by means of commonplace spreadsheet software or other computational software presumably bundled with these platforms (Hussain *et al.*, 2014).

(c) Students may even utilise such commonplace software to work out their answers, for example, in plotting graphs during online assessments of mathematical courses (Hussain *et al.*, 2014).

(d) Cumbersome input of mathematical notations and expressions into the computer platforms may impede students during online assessments (Anthony, Yang and Koedinger, 2005).

(e) The computer platforms' communication facilities may facilitate student cheating (Fask, Englander and Wang, 2014; Arnold, 2016; Dendir and Maxwell, 2020; Bilen and Matros, 2021), and mathematical courses typified by their absolute answers preclude most cheating accusation (Trenholm, 2007).

Moreover, even in the prior empirical researches that claimed to be on validity of online assessments in disciplines like psychology (Hewson, Charlton and Brosnan, 2007; Hewson, 2012) and medicine (Hope *et al.*, 2021), most of them simply compared students' online assessment mean scores to their offline assessment mean scores based on statistical tests of independent samples with scant regard for the consistency between online and offline scores in the same student sample. In other words, most such researches did not directly intend to verify convergent validity of online assessments.

The current study is to fill these gaps by evaluating the consistency between the online assessment scores (specifically, the invigilated, summative online assessment scores) and the offline assessment scores in the same student sample having taken a business mathematics course at the authors' university. State differently, the current study aims to verify the convergent validity of the invigilated, summative online assessment, i.e. the extent to which the invigilated, summative online assessment (as an instrument to measure the construct of the business mathematics course's learning outcomes) measures what is measured by an offline assessment broadly recognised as an instrument to measure the construct of the business mathematics course's learning outcomes. Further to convergent validity's verification, the current study additionally compares these two series of scores in order to deduce whether the students performed better in the invigilated, summative online assessment than in the offline assessment or vice versa. The comparison is based on statistical tests of paired samples as opposed to those of independent samples so as to compare the two assessment modalities with respect to the same student sample instead of, say, two randomly allocated samples.

2. THE STUDY

This study focused on evaluating the convergent validity of the invigilated, summative online assessment of a business mathematics course at the authors' university and the comparison between the online assessment scores and those of a corresponding offline assessment.

2.1 Materials and Methods

The course for this study concerned introductory business mathematics for year 1 students at the then School of Business and was delivered in the first semester of the academic year 2020/2021from September to January 2021. The total enrolments were 109. Instruction was offered offline and traditionally through classroom lectures but was complemented by the learning management system Canvas, which "tripled" as a course content repository (electronically storing all course materials), a course content retrieval system (for the students to retrieve all course materials) and an operational medium (for professor-student electronic communication, online assignment/test submission, mark announcement, etc.).

There was a mid-semester test conducted online through Canvas temporally around the middle of the semester and accounting for 40% of the course's overall assessment marks. This online test was the online assessment of which the convergent validity was to be verified in this study. It was in the form of a "quiz", which was a standard Canvas facility through which the professor could post the test questions and the students could answer by typing or uploading files. The files uploaded could be in the PDF, JPEG, PNG or Microsoft Word format and contain typed answers or scans/photographs of hand-written answers. Composed of "essay-type" mathematical questions asking for mathematical steps to arrive at mathematical solutions, the online test was invigilated (or proctored) and lasted one and a half hours in designated computer laboratories. It was graded/marked also online on Canvas afterwards partly automatically and partly manually by the professor in charge of the course, the marks of and the professor's comments for individual students also being announced to and accessible by the corresponding students on Canvas.

Towards the end of the semester, there was also an offline final examination, making up half the course's overall marks. Also comprising "essay-type" mathematical questions, the offline final examination was literally an offline assessment in that it was a traditional pen-and-paper examination with both questions from the professor printed and answers from the students hand-written on paper and was invigilated inside a physical hall/pavilion over a duration of three hours. Likewise, grading and commenting were performed on paper alongside students' answers manually by the professor in charge of the course. The topic coverage of the online test roughly constituted two-thirds of that of the offline final examination. In other words, only a third of the latter's topic coverage was on top of the former's. That is to say, the online test's contents resembled two-thirds of the offline final examination's, so the latter could act as a reference for the verification of the former's convergent validity.

2.2 Methods

By the close of the semester, the students' online test scores and their offline final examination scores became available, enabling the computation of the correlation coefficient between these two series of scores. A high coefficient would verify consistency between the online test scores and the offline final examination scores and to a large extent imply strong convergent validity of the online test in that the online test (as an instrument to measure the construct of the course's learning outcomes) measured what was measured by the offline final examination broadly recognised as an instrument to measure the construct of the course's learning outcome.

Then, a paired-sample *t*-test was performed to verify the hypothesis that students in the online test outperformed themselves in the offline final examination or vice versa. In the case of the hypothesis turning out to be statistically acceptable, one might be able to conclude that either of these two assessment modalities was to the students' benefit in terms of their scores.

3. RESULTS

Disregarding students absent from either the online test or the offline final examination, 99 out of the 109 students were qualified for the correlation analysis and the paired-sample *t*-test.

The correlation analysis gave rise to a correlation coefficient of 0.367 with a *p*-value of 0.000187, which was statistically significant even at the 1% significance level. With such a coefficient slightly below +0.5 and thus mildly on the low side, the implication was that the online test scores and the offline final examination scores were merely moderately consistent despite the substantial overlap between the coverage of the two assessments which were barely around two months apart. Stated differently, students scoring highly in the online test were only moderately likely to score highly in the offline final examination and vice versa. As such, the convergent validity of the online test, at least when gauged by the traditional, offline final examination, was just scarcely acceptable. Having said that, the *p*-value indicated that the correlation coefficient differed from zero even at the 1% statistical significance, attesting to the ability of the online test score being a proxy for the offline final examination score even if not a very faithful one.

As regards the paired sample *t*-test for the differences between the online test scores and the offline final examination scores for all the 99 students, the mean difference was 33.51 marks, its 95% confidence interval was from 28.189 to 38.832 marks, and the *p*-value of the *t*-test was 0.0000 and thus statistically significant even at the 1% significance level. The implication is that notwithstanding the appreciable overlap between coverage of the online test and the offline final examination which were marginally two months apart, students in the former outperforming themselves in the latter is almost undeniable, at least at the 1% significance level if put technically. On average, they scored 33.51 marks more highly in the former than in the latter. In other words, the online test favoured students in comparison with the offline final examination in that students scored drastically more highly in the former.

4. DISCUSSION

Prior literature does not accentuate the convergent validity of online assessments, or equivalently, the consistency between online assessments and traditional, offline assessments (or any alternative well-established and broadly recognised assessments). Instead most previous literature simply determines whether the mean scores of online assessments statistically differ from those of comparable offline assessments, presuming that online assessments are "fair and equitable" and worthy of pedagogical application as long as their mean scores do not deviate excessively from their offline counterparts (Hewson, Charlton and Brosnan, 2007; Hewson, 2012; Hope *et al.*, 2021). In fact, even if such differences between mean scores are negligible, there is no guarantee that online assessments truthfully reflect what offline assessments should reflect in respect of students' performance. For instance, even if there are zero mean score differences between online and offline assessments, it may be that eminent students ironically score lowly in online assessments but highly in offline assessments while less competent students happen to score conversely in the two assessment modalities, zeroing out the differences between the mean scores of the two. In contrast, even if the mean scores differ by a large margin, as long as online assessments render high scores for eminent students and low scores for less competent students as offline assessments supposedly do, the mean score disparity by no means prejudices online assessment's "fairness and equity" or debases online assessments one way or another at least when benchmarking against traditional, offline assessments. If optionally desired, the minor issue of mean score disparity can be muffled by simply scaling all students' scores across the board. In summary, it is the convergent validity of online assessments (or any assessment modalities else in question), or equivalently, the consistency between online assessments (or any assessment modalities else in question) and traditional, offline assessments (or any alternative well-established and broadly recognised assessments) instead of any mean score disparity between them that substantively determines the former's pedagogical worthiness.

This study exactly delved into the convergent validity of online assessments as exemplified by the scenario of a business mathematics course at the authors' university. It was found that such convergent validity was just scarcely acceptable. Additional findings were that students in the online assessment drastically outperformed themselves in the offline assessment, the two assessments being around two months apart.

Existing literature purports a students' computer anxiety and perception of and attitude towards online assessments to be exogenous/confounding variables/factors causing his/her online assessment score to deviate from his/her offline assessment score (Hewson, Charlton and Brosnan, 2007; Hewson, 2012). These variables/factors are considered exogenous/confounding in view of their not being any of the learning outcomes of the courses in question (unless in the currently irrelevant case of computer submersion courses whose learning outcomes may be computer proficiency, etc.) and thus are not supposed to be reflected in any assessment scores. Different students are prone to different levels of computer anxiety and possess different perceptions of and attitudes towards online assessments and thus are advantaged or disadvantaged differently in online assessments but not at all in offline assessments not involving computer platforms and/or online operations. Such anxiety, perceptions and attitudes might be some of the reasons for the different discrepancies between online assessment scores and offline assessment scores across different students and thus degraded convergent validity found in this study. Different students being advantaged or disadvantaged to different degrees in online assessments and thus degradation of online assessments' convergent validity in this study might have been strengthened by different students leveraging or being hindered by the five

distinctive features (a) to (e) of online assessments in mathematical courses (as enumerated in Section 1) to different extents. All such preliminary reasoning is to be investigated in further research.

By the same token, one may reason that the students in this study's online assessment drastically outperformed themselves in the offline assessment because students on average leveraged the five distinctive features (a) to (e) of online assessments in mathematical courses more than they were hindered. This reasoning is again subject to further research.

Besides, extension of this research to disciplines other than business mathematics for broader generalisation may be worthwhile just as what Pei and Wu (2019) hint.

5. CONCLUSION

Invigilated, summative online assessments for business mathematical disciplines are merely consistent with and can thus serve as proxies for invigilated, summative offline assessments to a limited extent. In addition, such online assessment scores substantially exceed their offline counterparts.

It is worth investigating whether computer anxiety and perception of and attitude towards online assessments, alongside distinctive features of mathematical courses' online assessment, are exogenous/confounding variables/factors underlying the disparity between online and offline assessment scores in this study and the convergent validity of online assessments falling short of impeccability. (Hewson, Charlton and Brosnan, 2007; Hewson, 2012).

REFERENCES

Anthony, L., Yang, J. and Koedinger, K. R. (2005) 'Evaluation of multimodal input for entering mathematical equations on the computer', in *Conference on Human Factors in Computing Systems - Proceedings*. Portland, OR: Association for Computing Machinery, pp. 1184–1187. doi: 10.1145/1056808.1056872.

Arnold, I. J. M. (2016) 'Cheating at online formative tests: Does it pay off?', *The Internet and Higher Education*, 29, pp. 98–106. doi: 10.1016/J.IHEDUC.2016.02.001.

Bilen, E. and Matros, A. (2021) 'Online cheating amid COVID-19', *Journal of Economic Behavior and Organization*. Elsevier B.V., 182, pp. 196–211. doi: 10.1016/j.jebo.2020.12.004.

Dendir, S. and Maxwell, R. S. (2020) 'Cheating in online courses: Evidence from online proctoring', *Computers in Human Behavior Reports*. Elsevier Ltd, 2. doi: 10.1016/j.chbr.2020.100033.

Dennick, R., Wilkinson, S. and Purcell, N. (2009) 'Online eAssessment: AMEE guide no. 39.', *Medical teacher*. England, 31(3), pp. 192–206. doi: 10.1080/01421590902792406.

Fask, A., Englander, F. and Wang, Z. (2014) 'Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment', *Journal of Academic Ethics*, 12(2), pp. 101–112. doi: 10.1007/s10805-014-9207-1.

Heale, R. and Twycross, A. (2015) 'Validity and reliability in quantitative studies', *Evidence-Based Nursing*, 18(3), pp. 66–67. doi: 10.1136/eb-2015-102129.

Hewson, C. (2012) 'Can online course-based assessment methods be fair and equitable? Relationships between students' preferences and performance within online and offline assessments', *Journal of Computer Assisted Learning*, 28(5), pp. 488–498. doi: 10.1111/j.1365-2729.2011.00473.x.

Hewson, C., Charlton, J. and Brosnan, M. (2007) 'Comparing Online and Offline Administration of Multiple Choice Question Assessments to Psychology Undergraduates: Do Assessment Modality or Computer Attitudes Influence Performance?', *Psychology Learning & Teaching*, 6(1), pp. 37–46. doi: 10.2304/plat.2007.6.1.37.

Hope, D. *et al.* (2021) 'Candidates undertaking (invigilated) assessment online show no differences in performance compared to those undertaking assessment offline', *Medical Teacher*. Taylor & Francis, 43(6), pp. 646–650. doi: 10.1080/0142159X.2021.1887467.

Hussain, M. A. *et al.* (2014) 'The Impact of Technology in Mathematics Education: A Case Study of Pakistan', *Journal of Basic & Applied Sciences*, 10, pp. 67–72. doi: 10.6000/1927-5129.2014.10.10.

Pei, L. and Wu, H. (2019) 'Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis', *Medical Education Online*. Taylor & Francis, 24(1). doi: 10.1080/10872981.2019.1666538.

Trenholm, S. (2007) 'A Review of Cheating in Fully Asynchronous Online Courses: A Math or Fact-Based Course Perspective', *Journal of Educational Technology Systems*, 35(3), pp. 281–300. doi: 10.2190/y78l-h21x-241n-7q02.

Whitelock, D. (2009) 'Editorial: e-assessment: developing new dialogues for the digital age', *British Journal of Educational Technology*, 40(2), pp. 199–202. doi: 10.1111/J.1467-8535.2008.00932.X.