

Open Research Online

The Open University's repository of research publications and other research outputs

An investigation of the validity of course-based online assessment methods: The role of computer-related attitudes and assessment mode preferences

Journal Item

How to cite:

Hewson, Claire and Charlton, John P. (2019). An investigation of the validity of course-based online assessment methods: The role of computer-related attitudes and assessment mode preferences. *Journal of Computer Assisted Learning*, 35(1) pp. 51–60.

For guidance on citations see [FAQs](#).

© 2018 John Wiley Sons Ltd



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.1111/jcal.12310>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

An investigation of the validity of course-based online assessment methods: The role of computer-related attitudes and assessment mode preferences

Claire Hewson, The Open University

John Charlton, Focus Statistics

To appear in Journal of Computer-Assisted Learning (pre-publication copy, August 2018).

Abstract

The use of e-assessment methods raises important concerns regarding the reliability and validity of these methods. Potential threats to validity include mode effects and the possible influence of computer-related attitudes. While numerous studies have now investigated the validity of online assessments in non-course-based contexts, few studies have addressed this issue in an educational context. The present study helps fill this research gap by investigating whether university students' computer-related attitudes and assessment mode preferences were related to performance on a course-based online assessment task. Overall, students' attitudes and preferences bore no greater relationships to performance on the online than offline module assessment tasks. This provides support for the validity of course-based online assessment methods and should help alleviate educators' concerns and encourage more widespread adoption of these methods, helping address the issue of their slow uptake to date. Suggestions for follow-up studies to corroborate and extend the current findings are offered.

Keywords: online assessment, validity, computer-related attitudes, computer anxiety, student preferences, assessment mode.

1. Introduction

Opportunities for adopting online assessment methods, across a range of contexts and applications, have recently greatly expanded due to the growth and penetration of the Internet and related technologies. Both researchers and educators have started to take up these opportunities and, accordingly, a body of research has started to emerge which investigates the use of these new forms of assessment, including their reliability and validity compared with more traditional offline methods. For example, a large number of studies have now been conducted to assess the validity of psychometric test instruments delivered online (e.g. Hewson & Charlton, 2005; Stephenson & Crete, 2010; Weigold, Weigold & Russell, 2013; van Ballegooijen et al, 2016; Yang et al, 2016). However, relatively few studies have directly considered the validity of course-based online assessment methods in an educational setting, despite the rapidly increasing interest in e- learning and assessment approaches over recent years (Havemann & Sherman, 2016; Soffer, Kahan & Livne, 2017). In particular, *online* delivery modes¹ have attracted much recent interest, given their capacity to respond to expanding student numbers, increasingly tight budgets, and student demands for convenience and flexibility. Various benefits of online assessment methods have by now been well-recognised, including scope for automated scoring and feedback (leading to both efficiency and reliability gains), enhancing student engagement and learning, and encouraging new, innovative approaches to assessment (Escudier et al., 2011; Farrell & Rushby, 2016; Hewson, 2012; Holmes, 2015; JISC, 2007). Within this context, a number of authors have recently identified the investigation of the validity and reliability of course-based online assessment methods as a current research need (Ardid et al., 2015; Farrell & Rushby, 2016; Fonolahi et al., 2014; Hewson, 2012; Meyer et al., 2016). The current lack of research on this topic may constitute a factor contributing to the relatively sluggish adoption of online assessment methods to date, particularly within the social sciences and humanities (Guàrdia, Crisp, & Alsina, 2016; Hewson, 2012; Warburton, 2009) and in 'high-stake' summative assessment contexts (Bennett et al., 2017; Boyle & Hutchinson, 2009; Clarke et al., 2004). Given this background, a

¹ While 'e-learning', more generally, refers to the use of ICT (Information and Communication Technologies) to support teaching and learning, *online* methods involve using the Internet, often in non-proctored contexts where students can access teaching and learning materials at a time and from a location of their own convenience.

primary goal of the present study was to help address this research gap by offering evidence on the validity of course-based online assessment methods. A secondary goal was to help stimulate additional research addressing this topic, so as to provide a sound evidence-base to help guide and inform practice and strategic decision making in this area, which currently is lacking (Whitelock, Gilbert, & Gale, 2013).

2. E-learning and the validity of course-based online assessment methods

One key question relating to e-learning approaches concerns whether students learning in the online mode fare as well as students engaged in traditional face-to-face (ftf) learning (Fonolahi, Khan, & Jokhan, 2014). For example, do students taking an equivalent course delivered in online and offline modes achieve equivalence of learning and equivalent overall pass rates and module grades? This is an important question, but not the key focus of the present paper. A related and equally important question – and the primary focus of the present study – is whether online assessment methods² which form an integral part of fully online course delivery can provide measures of student performance that are just as valid and reliable as offline assessment methods (such as traditional take-home essays). The validity of course-based online assessment methods is called into question if factors unrelated to student achievement of course-based learning outcomes interfere with performance on these tests. While it is common for students to now be able to submit traditional offline assessments, such as course-based essays and reports, via an online portal (e.g. as an uploaded word-processed electronic document), we reserve the term 'online assessment' for those assignments that require a more substantial level of interaction with a computer for their completion. It is in these interactive computer-based assessment contexts that it becomes plausible to suggest that mode and computer-related attitudes might impact on performance. It is our understanding that this is the standard usage of the terms 'online' and 'e-' assessment within the relevant body of literature. One possible threat to validity is unreliability of the technologies used to implement online assessments (Hewson, et al., 2007; Warburton, 2009). However,

² *Online* assessments are a form of *e-assessment*. E-assessment refers, more generally, to the use of ICT in assessment, whether online or offline, e.g., an in-class, proctored exam completed using a computer. Here, interest is primarily in *online* (e-)assessments using the Internet, often in non-proctored contexts (that is, taken outside of class at the student's convenience).

robust, well-tested, flexible systems have now been shown to offer reliable, user-friendly solutions for implementing online assessments, rendering technology reliability issues less of a concern today than previously. Thus, VLEs (Virtual Learning Environments) such as the well-established Blackboard (www.blackboard.com) and Moodle (<https://moodle.org>) platforms incorporate tools for authoring online assessments, offering solutions to support the effective integration of online assessment methods within online course delivery. Other factors which might pose a threat to the validity of online assessment methods are modality effects (online versus offline) and students' computer-related attitudes, such as computer anxiety and engagement, which may affect their performance in computerised testing situations (Beckers, Rikers, & Schmidt, 2006; Hewson, et al., 2007; JISC, 2007; Meyer et al., 2016; Powell, 2013). We now discuss each of these factors in turn, before describing the present study.

2.1 Online-offline modality effects and course-based assessments

In the context of the present discussion, modality effects occur when performance on an otherwise equivalent test or assessment is influenced by the mode, online or offline, in which the test is taken. Investigations of modality effects in a course-based assessment context have been lacking, compared with attention to this issue in non-course-based contexts (Fonolahi et al, 2014; Hewson, 2012). One notable exception directly addressing this question using an experimental design is presented by Hewson et al. (2007). They pseudo-randomly (based on surname) assigned undergraduate psychology students to take a low stake summative multiple choice question (MCQ) assignment either in online or pen and paper modes (both un-proctored). Both the online and offline groups received the assignment near the start of the module and were given the same amount of time to complete it. These authors found no effect of assessment modality on students' performance. More recently, Escudier et al. (2011) compared dental school students' performance on ftf and web-based administrations of a summative assignment, reporting that students performed equally well in both modes (both modes were offered in a proctored setting, as far as can be discerned). They used a repeated measures design, all students completing both the 'online' (proctored computerised) and offline modes, which were generated

by splitting the test in half and administering one half via a computer and the other half in pen and paper format, counterbalanced so that one group received the computerised test items first while the other group received the pen and paper items first. These authors concluded that online assessment does not seem to confer any overall disadvantage to students. However, the proctored nature of the assessments should be noted here, whereas the present study is particularly concerned with unproctored online contexts, such as was used by Hewson et al. (2007), which have been under-studied to date. Other authors have also noted a lack of modality effects when comparing proctored, computerised course-based assessments with pen and paper administrations (e.g., Ashton et al., 2005; Cassady & Gridley, 2005). Kingston (2009) reviewed 81 studies (all of which seemed to have used a proctored in-class setting), comparing computer- and paper-administered MCQ tests, in schools rather than higher education settings, and concluded that only very small effect sizes have been observed in these studies overall. Meyer et al. (2016) report a lack of modality effects when using a more complex form of online assessment; they found that students performed equivalently in online and offline (traditional 'laboratory') versions of an anatomy exam, concluding that their findings will allow anatomists to confidently implement online assessments without fear of jeopardising academic rigour or student performance. Similarly to other studies mentioned here, the 'online' version of the assessment used in the study of Meyer et al. (2016) was actually a proctored computerised test (taken in-class under supervision), rather than a 'truly online' unproctored assessment.

Some studies have found evidence for modality effects, reporting poorer performance in computerised than in pen and paper testing modes (Goldberg & Pedulla, 2002; Ricketts & Wilks, 2002). These studies have tended to use proctored, in-class computerised contexts, rather than truly online settings in which students complete the test at a time and place of their own convenience (Hewson, 2012). Also, the presence of factors confounded with modality (such as using non-equivalent tests, or non-equivalent learning environments, between the two modality groups) often precludes drawing any firm conclusions without follow-up research (Hewson, 2012). It should also be noted that these studies are now quite dated. Further studies investigating this topic would therefore be of value, particularly in the context of

fully-online courses as opposed to blended learning contexts which use a combination of online/computerised and face-to-face learning and assessment, since the former have received very little attention to date (Stödberg, 2012).

Closer attention to possible relationships between course delivery mode and course assessment mode would also be helpful, since there is some evidence that these may interact to influence assessment performance. Fonolahi et al. (2014) found that while students taking a module in offline (ftf) or online mode performed equivalently in terms of overall module pass rates, those students taking the module online performed better on coursework and less well on the exam, relative to the offline group. Since the online group completed the coursework assessments online (using Moodle) whereas the offline (ftf) group completed them in pen and paper format (hence confounding assessment modality with course delivery mode), and since both groups took online formative practice tests, one possibility is that the online group performed better in the coursework due to having had practice in this testing mode. In contrast, the offline group had no such practice in the offline assessment mode since there were no equivalent pen and paper practice tests. In any case, given that the assessments for the two groups were not equivalent in terms of content, the move to online course delivery having occurred subsequent to, and fully replacing, the offline mode of delivery, follow-up research is needed before any firm conclusions can be drawn regarding the findings from this study. This illustrates the limitations of practice-based evaluations, compared with tightly controlled experimental designs (which can be challenging to implement in naturalistic educational settings). Nevertheless, the study of Fonolahi et al highlights an important consideration relevant to the question of the validity of online assessment methods: namely, performance in these new assessment delivery modes may also be related to the mode in which a course is delivered. Research on this topic should therefore take this consideration into account. Conversely, when investigating whether students taking a module online or offline show equivalence of learning, the modality of the testing environment should also be taken into account since it should not be assumed that assessment modality effects will not occur when non-equivalent modalities (computerised/online and pen and paper) are used.

Overall, the evidence to date seems to point to a lack of modality effects when comparing computerised/online and pen and paper assessments in a course-based setting. However, some of the findings have been equivocal and require follow-up studies. Particularly, further research is needed which considers modality effects in relation to un-proctored (online/offline) assessment tasks, and which implements tighter control over possible confounds with modality (following the example of Hewson et al. 2007).

2.2 Computer-related attitudes and computerised task performance

The relationship between computer attitudes, including anxiety, and computer-mediated behaviour continues to raise important issues in a range of settings, including leisure, education and the workplace (e.g. Cazan, Cocoradă & Maican, 2016; Celik & Yesilvurt, 2013; Osatuvi, 2015). While the association between performance on computer-based tasks and computer-related attitudes has been widely studied (Brosnan, 2002; Powell, 2013), there remains a lack of research exploring the possible influence of computer-related attitudes in a *course-based* assessment context (Hewson et al., 2007; Hewson, 2012), and the limited findings on this topic to date have been equivocal (Mahar, Henderson, & Deane, 1997; Beckers et al., 2006; Powell, 2013). The educational context remains distinctly relevant because of the increased prevalence of e-learning and the implications for professional attainment. While students' computer- and technology-related attitudes, skills and behaviours will no doubt have changed in various ways over recent years, this does not preclude the possibility that some students may remain disadvantaged by contexts which require computerised teaching-learning modes to be used, including e-assessments. There is, in particular, a notable lack of research which considers the role of computer-related attitudes in an *online* (un-proctored) course-based assessment context (Hewson, 2012). The drive to introduce e-assessment methods in education presupposes that candidates are competent with, and willing to engage with, ICT (JISC, 2007). However, 'for some learners, this could create a new layer of disadvantage, if they feel unable or unwilling to participate in courses and assessments that take place in a computer-mediated environment' (JISC, 2007, p.37). A key goal of the present study was to contribute in addressing this concern by exploring relationships between

computer-related attitudes and performance on an online course-based assignment. To our knowledge, only one previous study has directly addressed this issue: Hewson et al. (2007) considered relationships between computer anxiety, computer engagement, and performance on a simple, course-based, low-stake summative online MCQ assignment used to assess knowledge gained from a module delivered using traditional ftf teaching/learning methods (lectures and seminars). They found little evidence that students' levels of computer anxiety and computer engagement were related to performance on the online assignment. The current study set out to investigate whether these findings could be replicated in a different setting, extending the work of Hewson et al., (2007) by using a more complex and higher stake summative computerised assessment, a non-traditional student group (mature, part-time), and a fully-online/distance course.

In addition to more general computer-related attitudes such as computer anxiety and computer engagement, students' attitudes towards online assessment methods themselves are relevant to questions about the validity and fairness of e- and online assessment approaches. Several studies have investigated students' attitudes towards e-/online assessment methods, some finding these to be generally positive (Escudier et al., 2011; Marriott, 2009; Meyer et al., 2016), others reporting less positive attitudes (Hewson et al., 2007), and some studies indicating a shift from less positive to more positive attitudes over time (Holmes, 2015). Studies investigating how such attitudes may be related to actual performance in online and offline assessment modes, however, remain sparse. Hewson (2012) investigated this issue, reporting that students' self-reported preference for taking an assignment in either online or offline mode was not related to their performance in either mode (students were assigned to an assessment modality, so were not able to self-select to use their preferred mode). Meyer et al. (2016) investigated whether students' perceptions of laboratory and online testing environments were related to their performance in either of these modes; they report finding no relationships between perceptions and performance for either the offline or online mode. These findings suggest that students with less favourable attitudes toward online assessments are not disadvantaged by having to use this assessment mode, at least not in terms of performance scores achieved. Of course, students may still be disadvantaged in terms of the quality of the student experience if

they are required to use an assessment mode which leads them to feel worried or anxious (Hewson, 2012; JISC, 2007). Further investigations into students' attitudes towards online assessment modes, and how these may relate to performance in these modes, would be of value. The present study makes a contribution by investigating this issue.

3. The present study

As noted above, investigations of factors which might pose a threat to the validity of (course-based) e-assessment methods – particularly mode and computer-related attitudes – have been limited to date. This is especially the case for *online* assessments, i.e. those administered via the internet in un-proctored contexts (Hewson, 2012). Also, it has been pointed out that there is a particular need for research in the context of *fully online* courses (Stödberg, 2012; Fonolahi et al., 2014), and contexts using more complex and (particularly higher-stake) summative online assessments (Hewson et al., 2007; Hewson, 2012). The present study addresses these needs by investigating the validity of a summative, course-based assessment for a fully online/distance social science Masters in Research Methods module. The online assessment used involved interacting with an SPSS (Statistical Package for the Social Sciences) data set to derive answers so as to be able to answer MCQ questions and questions in other formats (e.g., drag and drop). In addition, the current study extends existing work by using a non-traditional student participant group, consisting primarily of part-time, mature students.

The first main research question considered whether there were any relationships between students' computer-related attitudes (computer anxiety and computer engagement) and their performance on the online and offline module assessments – more specifically, whether any observed relationships were greater for the online than the offline assessments (all students completed three module assignments, one online and two offline). The second main research question asked whether students' mode-preference for taking an assessment (online, offline or no strong preference) was related to their performance on the online assignment. Other questions of interest concerned whether there were any relationships between demographic variables (sex and age) and the computer-related attitudes and module assessment performance

scores. Some previous studies have found sex and age to be related to computer anxiety (females and older users having higher levels of anxiety), though, overall, the existing literature is equivocal on this issue (Cazan et al., 2016; Fernández-Ardèvol & Ivan, 2015; Fuller et al., 2006; Lee & Huang, 2014; Powell, 2013).

3.1 Method

3.1.1. Participants

Eighty-nine students studying the masters-level, social science module *D849 Introduction to Quantitative and Qualitative Research Methods* at the Open University participated (26 males, 63 females; age range 24-84, $M = 43.83$ years, $SD = 11.40$ years). The sample was recruited from three cohorts, each studying the module during different years that it was run. Participants were volunteers, recruited via a participation request message sent directly to their university email account.

3.1.2 Materials

Materials consisted of an online assessment (*iCMA*: interactive, computer-marked assessment), a computer attitudes questionnaire, and some additional questions to probe attitudes towards online assessments. The *iCMA*, implemented using Moodle, was compulsory for all students taking the *D849* module, and constituted 30% of the overall marks for the continuous assessment component, which itself made up 50% of the total module mark. The remaining 70% of marks for continuous assessment were gained by completing an essay-style question (*TMA*: tutor-marked assignment). An end-of-module assessment (*EMA*, consisting of a multi-part essay/report assignment) made up the other 50% of the total module mark. The *iCMA* tested students' understanding of quantitative research methods and SPSS skills as taught within the first seven weeks of the module. The *iCMA* contained a mixture of multiple choice and other types of question format (e.g., interactive 'drag and drop' questions), and required students to download and work with an SPSS data file to obtain answers to some questions (ten out of twenty questions included in the *iCMA*). The *iCMA* was scored automatically by the computer, removing any scope for human error, and for each question presented the possible answer choices from which students were able to select were unambiguously either 'correct' or 'incorrect', resulting in this being a

highly reliable assessment measure. The final iCMA assignment mark consisted in a score out of 100.

The TMA assignment was completed near the mid-point of the module, and marked by a tutor. The mark obtained consisted in a score out of 100. The EMA assignment was completed near the end of the module, and was marked by a tutor. The mark obtained consisted in a score out of 100. For both the TMA and EMA assignments, rigorous quality control procedures were utilised to ensure and assess the reliability of these measures. These included: providing all marking tutors with a detailed marking guide; requiring marking tutors to engage in a group marking coordination exercise; monitoring of a sample of scripts for each marking tutor; monitoring of a sample of TMA and EMA scripts by an external examiner; comparing performance of students taking the module in different years of presentation on the TMA and EMA assignment measures (see above). The outcome of these processes allowed us have a high level of trust in the reliability of the EMA and TMA measures we used in the present study.

The computer attitudes questionnaire was as described in detail in Hewson et al. (2007), and consisted of a computer anxiety (CA) subscale (17 items, a higher score indicating higher anxiety, with minimum and maximum possible scores of 17 and 85) and a computer engagement (CE) subscale (12 items, a higher score indicating higher engagement, with minimum and maximum possible scores of 12 and 60), derived from factors in the factor analysis of Charlton (2002). Students responded to items that were statements concerning their computer-related thoughts and behaviours, using a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*. Cronbach's alpha coefficients for the present data set showed that both subscales were reliable, computer anxiety $\alpha = .89$, and computer engagement $\alpha = .79$.

Additional questions presented to students after they had completed all questionnaire items included: a) how often they use a computer (daily, two or three times a week, weekly, less than weekly) and; b) how they would prefer to take a module assignment (online, offline (pen and paper), no strong preference).

3.1.3. Design and Procedure

Students were required to complete the iCMA during a five week period from when it first became available, via the module website, during Week 5 of the module, and when it closed in Week 10. Students were free to navigate forwards and backwards through the iCMA questions during completion, and could save their answers to revisit at a later time until they submitted their final answers by the assignment deadline.

The computer attitudes questionnaire, including the additional two questions appended at the end and some initial demographic questions (sex and age), was completed after having completed the module (after students had completed all assignments, including the final EMA assessment). This questionnaire was administered and completed online using SurveyMonkey (see surveymonkey.com).

4. Results

Preliminary analyses revealed no significant differences for any of the variables analysed between the three different student cohorts who had taken the module across subsequent years that it was run. Therefore data were pooled for the cohorts in the analyses reported below. In relation to the module assessment measures, this finding was to be expected, on the assumption that student cohorts across subsequent years are comparable, and that the assessments used across subsequent years are reliable measures of student achievement. In relation to the computer attitudes measures, this indicates that students taking the module from year to year did not differ significantly in their computer-related attitudes.

To provide an overview of sample characteristics in addition to the information on demographic characteristics of the sample provided in the Method section, it is useful to observe that in response to the question asking about students' preferred mode of assignment completion, a large majority (75.8%, $n = 69$) preferred online completion, only 3.3% ($n = 3$) preferred offline completion, and 20.9% ($n = 19$) had no strong preference. In response to the question asking how often they currently used a

computer, nearly all students (97.8%, $n = 89$) used a computer every day, 2.2 % used a computer two or three times per week, and no students used a computer only once per week or less.

4.1 Computer anxiety, computer engagement and performance

Five cases were identified as having outlying values ($z > +/- 3.00$) across the attitudinal and performance variables (EMA mark and iCMA mark each had two cases with negative values, and computer anxiety had one case with a positive value). These cases were excluded from all analyses, including the reporting of the descriptive statistics for the attitudinal and performance variables in Table 1.

Table 1

Descriptive statistics for attitudinal and performance variables after deletion of outliers

	<i>M</i>	<i>SD</i>	<i>n</i>
Computer anxiety	34.39	8.76	84
Computer engagement	40.98	5.75	83
iCMA mark	79.99	13.28	80
TMA mark	68.06	13.73	78
EMA mark	66.58	15.23	78

In an initial analysis it was of interest to compare the computer anxiety and engagement scores of the present sample with those of the samples analysed by Hewson et al. (2007)³, with a view to considering whether any observed differences in

³ Access to the data set of Hewson et al. (2007) was possible since that study was a previous study carried out by the present authors. These data are not available for public access.

attitudes might be explained either by changing attitudes over time, or sample differences (see discussion for further comments on this point). Here, independent samples *t*-tests showed that the present sample were significantly lower in anxiety ($M = 34.39$, $SD = 8.76$, $n = 84$) and significantly higher in engagement ($M = 40.98$, $SD = 5.75$, $n = 83$) than the combined 2007 samples ($M = 39.91$, $SD = 10.45$, $n = 97$ for anxiety and $M = 36.85$, $SD = 5.92$, $n = 97$ for engagement), $t(179) = 3.80$ for anxiety ($p < .001$, two-tailed) and $t(178) = 4.73$ for engagement ($p < .001$, two-tailed). With respect to effect size, both analyses exceeded Cohen's (1988) benchmark of $d = .5$ for a medium effect, and that for engagement approached the benchmark of $d = .8$ for a large effect: $d = .57$ for anxiety and $d = .71$ for engagement. To test the first main research question asking whether computer-related attitudes were more highly related to iCMA mark than scores on the two offline performance variables, a Pearson's r correlation matrix was produced with listwise deletion of cases for ease of comparison of coefficients (see Table 2). From the matrix it can be seen that there was a significant small to medium negative relationship in which iCMA mark decreased as computer anxiety increased (Cohen's 1988 benchmarks being $r = .10$ for a small effect and $r = .30$ for a medium effect). Also, while the positive coefficient indicating an increase in iCMA mark with increasing computer engagement exceeded Cohen's benchmark for a small effect size, this coefficient did not reach significance. The only relationship between the computer attitude variables and the two offline academic performance measures was one whereby increasing computer engagement was related to increasing TMA essay marks, this relationship approaching a medium effect size.

Table 2

Correlations (Pearson's r) between the computer attitudes and performance variables ($N = 72$)

	Computer Anxiety	Computer Engagement	iCMA	TMA essay
Computer anxiety	-----			
Computer engagement	-.62**	-----		
iCMA mark	-.26*	.18	-----	
TMA essay mark	-.12	.26*	.28*	
EMA report mark	-.01	.05	.36**	.37**

* $p < .05$, ** $p < .01$ – one-tailed

A test of the difference in the size of the computer anxiety – iCMA mark and computer anxiety – TMA essay mark relationships revealed no significant difference, $z = -1.00$, $p = .16$, one-tailed. While perhaps unnecessary given the negligible computer anxiety – EMA report mark relationship, a test of the difference in the size of the computer anxiety – iCMA mark and computer anxiety – EMA report mark relationships confirmed that the former coefficient was significantly larger, $z = 1.87$, $p = .03$, one-tailed. Note that although iCMA mark was significantly correlated with both TMA essay and EMA report marks, the latter two offline performance variables were not significantly correlated with computer anxiety, and therefore it was not surprising to find that analyses in which first TMA essay mark and then EMA report mark were partialled out of the computer anxiety – iCMA mark relationship confirmed that controlling for the two offline performance measures had little effect on the relationship between computer anxiety and iCMA mark; with TMA essay mark partialled out $r_{12.3}(71) = -.24$, $p = .04$, two-tailed, and with EMA report mark partialled out $r_{12.3}(71) = -.28$, $p = .02$, two-tailed.

With respect to the research question asking whether computer-related attitudes were more highly related to iCMA mark than scores on the two offline performance

variables, the above analyses showed that iCMA mark decreased as anxiety increased, and that computer anxiety’s relationship with iCMA mark was greater than that for its relationship with EMA report mark but not TMA essay mark. On the other hand, computer engagement was not significantly related to iCMA scores, but it was significantly related to TMA essay mark, although the difference in the size of relationships was not significant, $z = 0.57, p = .28$, one-tailed.

4.2 Assessment mode preferences, iCMA mark and computer attitudes

The second main research question asked whether there was a difference in performance on the online assessment (iCMA) according to students’ preferred mode of assignment completion. Since only three participants expressed a preference for offline assessment inferential analysis relating to this research question took the form of an independent samples *t*-test comparing students declaring a preference for online assignment completion with those saying they had no strong preference. To put this result in context, *t*-tests were also performed for the two computer attitude measures. Descriptive statistics for these tests can be seen in Table 3.

Table 3

Descriptive statistics for independent samples *t*-tests examining differences in iCMA scores and computer attitude variables according to assessment modality preference

	Modality preference					
	Online			No strong preference		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
iCMA mark	79.53	13.85	62	83.40	10.43	15
Computer anxiety	33.00	7.49	64	37.35	11.21	17
Computer engagement	41.67	5.43	64	38.63	6.15	16

The test for iCMA marks revealed no significant difference between the two aforementioned groups, $t(75) = 1.01, p = .32$, two-tailed. However, the effect size ($d = .32$) was in between Cohen's benchmarks for small and medium effects ($d = .2$ and $.5$ respectively). So, with respect to the research question, although results did not achieve significance, students with a preference for online assessment actually performed slightly worse in the online assessment than those with no such preference (note also that two similar t -tests for the two offline assessments revealed no significant differences between the two groups, $p = .96$ and $p = .39$ – two-tailed for TMA essay marks and EMA report marks respectively). Finally, for one-tailed tests at least, as would be expected, there were significant differences in computer anxiety, $t(79) = 1.90, p = .03, d = .46$, and computer engagement, $t(78) = 1.95, p = .03, d = .52$, with anxiety being significantly lower and engagement being significantly higher for the group with a preference for online assessments, and with medium effect sizes being observed.

4.3 The potential influence of demographic factors

Pearson's r analysis was also used to test whether age exhibited any relationships with computer anxiety, engagement and iCMA mark. This showed that no such relationships existed: for computer anxiety, $r(77) = .06, p = .58$, two-tailed; for computer engagement, $r(76) = .05, p = .64$, two-tailed; for iCMA mark, $r(74) = .03, p = .77$, two-tailed. None of these correlations exceeded Cohen's (1988) benchmark for a small effect size. No significant relationships were observed between age and either of the two offline module assignments.

Independent samples t -tests were used to test for sex differences in computer anxiety, engagement and iCMA mark. Table 4, which gives the descriptive statistics for these tests, shows that there were only minor differences in means for males and females, and this impression was confirmed by the results of the inferential tests, which revealed that no sex differences existed: for computer anxiety, $t(79) = 0.97, p = .33$, two-tailed, $d = .23$; for computer engagement, $t(79) = 0.52, p = .61$, two-tailed, $d = .12$; for iCMA mark, $t(76) = -0.33, p = .75$, two-tailed, $d = -.08$. Note that the effect

for computer anxiety just exceeded Cohen’s (1988) benchmark for a small effect size ($d = .2$), though the observed sex difference was not significant. No significant relationships were observed between sex and either of the two offline module assignments.

Table 4

Descriptive statistics for independent samples *t*-tests examining sex differences in computer anxiety, computer engagement and iCMA scores

	Females			Males		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Computer anxiety	34.76	8.40	58	32.70	9.41	23
Computer engagement	41.29	5.44	58	40.57	6.45	23
iCMA mark	79.80	13.53	56	80.91	13.40	22

5. Discussion

This study set out to address the issue of the validity of online course-based assessment methods, an issue identified as constituting a research gap in the existing e-assessment literature. Two key research questions were addressed: first, whether there was any relationship between students' computer-related attitudes (computer anxiety and computer engagement) and their performance on online and offline module assessments, in particular whether any observed relationships were greater for the online than the offline assessments; second, whether students' mode-preference for taking an assessment (online, offline or no strong preference) was related to their

performance on an online assignment. In addressing these questions the study also extended the generalisability of the scant existing research on this topic by using a higher-stake and more complex type of online assignment than has been studied previously, completed by a non-traditional student sample (mature, part-time) enrolled on a fully-online/distance module, all of which have been identified as existing research needs (Hewson et al., 2007; Hewson, 2012; Fonolahi et al., 2014). Considering the first main research question, it was found that, overall, students' attitudes (computer anxiety and computer engagement) showed no greater relationships with the online than offline assignment marks. This suggests that students with differing computer-related attitudes were not differentially (dis)advantaged by the online assessment modality. This finding supports the scant work on this topic to date (Hewson et al., 2007) and is important in offering further support for the validity of course-based online assessment methods and encouraging their more widespread adoption. In relation to the second main research question, which considered whether students' self-reported preferences for taking an assignment online or offline were related to online assignment marks, our findings revealed no such relationships. This corroborates previous work that has considered whether students' preferences (Hewson, 2012) and perceptions (Meyer et al., 2016) regarding online and offline assessments may be linked to their performance in these modes. This finding offers additional support for the validity and adoption of online assessment methods. Finally, possible associations between sex, age and computer-related attitudes were considered in the present analysis, since previous research has found these demographic variables to be related to computer attitudes (Powell, 2013). Computer anxiety and computer engagement were found not to be related to sex or age, although a non-significant small effect was observed such that females showed slightly higher mean computer anxiety scores than males, which is consistent with previous literature (Powell, 2013). Sex and age were not related to performance on either the online or offline module assignments. On the whole, the present findings are encouraging in relation to the overarching research question of interest here which is whether course-based online assessment methods can offer a valid and fair measure of student ability. Nevertheless, some of the present findings warrant further discussion and point toward useful avenues for follow-up research, as now discussed.

5.1 Future research directions

It is noteworthy that, despite computer attitudes showing no greater relationships with online than offline assignment marks, overall, some significant small to medium effect relationships were observed between computer-related attitudes and assignment marks. Thus computer anxiety was negatively related to the online assignment mark ($r = -.26, p < .05$), indicating that higher levels of anxiety were associated with lower marks, and computer engagement was positively related to the offline essay assignment mark ($r = .26, p < .05$), indicating that higher levels of engagement were associated with higher marks. None of the other small to medium correlations observed were significant, but all were also in the direction to be expected if more positive computer attitudes are associated with better performance. This indicates that computer-related attitudes may play some role in how students fare in e-learning contexts, though as already noted the present findings can not be explained as a straightforward impact on online assignment performance since computer attitudes were also found to be related to the offline assignment marks. Referring back to the earlier discussion regarding whether students achieve equivalence of learning when taking a course in online and offline modes (Fonolahi et al., 2014), one possibility is that students' computer-related attitudes (e.g. computer anxiety) might impact upon the extent to which they are willing to engage with e-learning approaches: lower levels of engagement could potentially lead to reduced learning in online course delivery contexts, and subsequent worse performance on course-related assignments, thus disadvantaging some students (Fuller et al., 2006). There is some evidence that students with less positive computer-related attitudes do engage less with e-learning environments and tools (Fuller et al., 2006; Powell, 2013). While the present study did not directly measure students' engagement with the online teaching/learning materials, the observation that computer engagement was positively related to performance on the offline essay assignment is suggestive that students with higher levels of computer engagement may have achieved better learning in the online teaching-learning environment. Regarding the observed negative relationship between computer anxiety and performance on the online assignment, this was not significantly greater than the similar relationship observed between computer anxiety

and performance on the *offline* essay assignment, suggesting that computer anxiety may, at least in part, have impacted upon learning success within the online teaching-learning environment. It should also be noted, however, that only negligible correlations were observed between both computer anxiety and computer engagement and the offline end-of-module assignment (EMA). Clearly, future studies that further interrogate these potential relationships between computer-related attitudes and course-based delivery and assessment modes would be helpful in clarifying and extending the present findings. In particular, studies that directly measure students' levels of engagement with e-learning materials (which could be readily achieved using current data analytics approaches) and explore how these may be related to computer-related attitudes would be informative, as would studies that directly compare levels of engagement and learning success within online and offline course delivery modes. Studies that directly tackle the issue of how teaching-learning delivery mode (online, offline) and assessment mode (online, offline) may interact, as discussed earlier, are also of clear value, and may have potentially important implications for e-learning/assessment design.

Future studies should also explore the generalisability of existing research on the validity of online assessments to other e-learning/assessment contexts, including investigating alternative types of assessment task and different student groups. The present study extended the work of Hewson et al. (2007) by using a more complex and higher-stake online assessment task, a non-traditional student group and a fully-online/distance course. Whilst reaching the same overall conclusions regarding supporting the validity of online assessment methods compared with traditional (offline) approaches, the two studies differed in that whereas the present study observed some significant relationships between computer-related attitudes and (online and offline) assessment performance, Hewson et al. (2007) observed no such relationships. It is not clear exactly which factors might best account for these differences (course delivery mode would seem to be a prime candidate, in light of the immediately preceding discussion), and further studies would help offer valuable additional insights into the roles each factor might play. Computer experience is an individual difference factor worthy of further investigation, particularly since previous research has found computer experience to be negatively related to computer

anxiety (Powell, 2013; Saadé & Kira, 2009). The participants in the present study had high levels of computer experience and relatively low levels of computer anxiety, which is consistent with this prior research. Studies exploring the issues raised here with lower-experience groups would be of value. Prior research has also found computer experience to mediate the relationship between sex and computer anxiety (Powell, 2013), so studies investigating whether sex differences may exist in lower-experience groups are needed (and similarly for age). Future research could also usefully explore the role of other computer-related attitudes, beyond those considered in the present study, within an e-learning context. For example computer self-efficacy (CSE) has been found to be strongly related to computer anxiety (Saadé & Kira, 2009) as well as to moderate the relationship between computer anxiety and performance on a computer-related task (Saadé and Kira, 2009). Strategies for raising students' levels of CSE could be built into e-learning design should this factor turn out to be important when considering students' performance in this context. While there is little doubt that students' use, skills and attitudes towards teaching-learning technologies, including online course delivery and assessment methods, will have undergone various changes over recent years, there is an existing need for more up-to-date research that considers the nature of these changes, as well as the extent to which technology-enhanced teaching-learning methods can offer reliable, valid solutions that equally benefit all types of student. The present study reinforces this need by indicating that while the online assignment examined here did not seem to disadvantage any of the students who took part in the present study, some small significant relationships were observed between students' computer-related attitudes and their performance on some of the (offline and online) assessment tasks. We have tentatively suggested that this might be explained by students with differing computer-related attitudes engaging differently within an online/distance learning environment. Further research on this topic would be of value.

Finally, further research into students' e-learning and e-assessment perceptions and preferences would be of value. While the present study found these preferences not to be related to performance, supporting prior research (Hewson, 2012; Meyer et al., 2016), most students in the present study expressed a preference for online assessment

(76%) with very few preferring offline assessments (3%), in line with previous studies (Marriott, 2009; Sheader, Gouldsborough & Grady, 2006). The fact that the findings reported here contrast with those in Hewson (2012), who found lower levels of preference for online assessments, suggests a trend whereby as students are becoming increasingly familiar and comfortable with e-assessment approaches, perceptions and preferences are likely to be changing (see also Holmes 2015). It should also be noted, however, that the students who took part in the present study had chosen to enrol on an online distance learning course, and preferences may differ for other groups.

6. Conclusion

The present study has contributed to the limited research to date on the issue of the validity of online assessment methods, particularly the possible role of computer-related attitudes in relation to performance on these still relatively novel forms of assessment. This study extended existing work by investigating these issues, a) using a higher-stake summative, more complex form of online assessment task than has been examined previously, and, b) using a non-traditional (primarily part-time, mature) student group taking a fully online/distance module. Despite these differences, there was little evidence to suggest that taking an assignment in the online assessment modality disadvantaged students with less positive computer-related attitudes. Also, students' self-reported preferences for taking an assignment online or offline were found not to be related to online assignment performance scores. Overall, these findings are encouraging in relation to the question of whether online assessments can offer a valid, reliable, fair measure of student attainment in terms of having met course-related learning outcomes. Nevertheless, the observation of some small to medium relationships between computer-related attitudes and module assignment marks (relating to both online and offline assignments) indicates a need for further research in this area. It is possible that some students may be disadvantaged by e-/online teaching and learning approaches because of their relatively low level engagement with computers. Future studies may usefully explore these issues using different types of assessment, different user groups, and different teaching/learning environments. Several suggestions for fruitful research avenues have been offered here. It is hoped that these suggestions may contribute toward the

development of a rigorous, robust, reliable evidence-base for e-assessment practice, that has been lacking (Whitelock et al., 2013).

6. References

Ardid, M., Gómez-Tejedor, J. A., Meseguer-Dueñas, J. M., Riera, J., & Vidaurre, A. (2015). Online exams for blended assessment. Study of different application methodologies. *Computers & Education*, *81*, 296 – 303.

<https://doi.org/10.1016/j.compedu.2014.10.010> (last accessed 21st Nov, 2017).

Ashton H.S., Beevers C.E., Korabinski A.A., & Youngson M.A. (2005). Investigating the medium effect in school chemistry and college computing national examinations. *British Journal of Educational Technology*, *36*(5), 771-787. doi: 10.1111/j.1467-8535.2005.00501.x.

Beckers, J. J., Rikers, R. M., & Schmidt, H. G. (2006). The influence of computer anxiety on experienced computer users while performing complex computer tasks. *Computers in Human Behavior*, *22*(3), 456-466. doi: 10.1016/j.chb.2004.09.011.

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology*, *48*(2), 672-682. doi: 10.1111/bjet.12439.

Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: What are they and what are their benefits? *Assessment & Evaluation in Higher Education*, *34*(3), 305-319. doi: 10.1080/02602930801956034.

Brosnan, M. J. (2002). *Technophobia: The psychological impact of information technology*. Routledge.

Cassady, J.C., & Gridley, B.E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning & Assessment*, *4*(1-30). Available at:

<http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1648/1490> (last accessed 21st Nov, 2017).

Cazan, A. M., Cocoradă, E., & Maican, C. I. (2016). Computer anxiety and attitudes towards the computer and the internet with Romanian high-school and university students. *Computers in Human Behavior*, 55, 258-267. doi: 10.1016/j.chb.2015.09.001.

Celik, V., & Yesilyurt, E. (2013). Attitudes to technology, perceived computer self-efficacy and computer anxiety as predictors of computer supported education. *Computers & Education*, 60(1), 148-158. doi: 10.1016/j.compedu.2012.06.008.

Charlton, J.P. (2002). A factor analytic investigation of computer addiction and engagement. *British Journal of Psychology*, 93(3), 329-344.

Clarke S., Lindsay K., McKenna C., & New, S. (2004) INQUIRE: A case study in evaluating the potential of online MCQ tests in a discursive subject. *ALT-J, Research in Learning Technology*, 12, 249–260. doi: 10.1080/0968776042000259564.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Escudier, M. P., Newton, T. J., Cox, M. J., Reynolds, P. A., & Odell, E. W. (2011). University students' attainment and perceptions of computer delivered assessment: A comparison between computer-based and traditional tests in a "high-stakes" examination. *Journal of Computer Assisted Learning*, 27(5), 440–447. doi: 10.1111/j.1365-2729.2011.00409.x.

Farrell, T., & Rushby, N. (2016). Assessment and learning technologies: An overview. *British Journal of Educational Technology*, 47(1), 106-120. doi: 10.1111/bjet.12348.

Fernández-Ardèvol, M., & Ivan, L. (2015, August). Why age is not that important? An ageing perspective on computer anxiety. In *International Conference on Human Aspects of IT for the Aged Population* (pp. 189-200). Springer, Cham.

Fonolahi, A. V., Khan, M. G. M., & Jokhan, A. (2014). Are students studying in the online mode faring as well as students studying in the face-to-face mode? Has equivalence in learning been achieved? *Journal of Online Learning and Teaching*, 10(4), 598.

Fuller, R. M., Vician, C., & Brown, S. A. (2006). E-learning and individual characteristics: The role of computer anxiety and communication apprehension. *Journal of Computer Information Systems*, 46(4), 103-115.

Goldberg A.L., & Pedulla J.J. (2002). Performance differences according to test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement*, 62, 1053–1067. doi: 10.1177/0013164402238092.

Guàrdia, L., Crisp, G., & Alsina, I. (2016). Trends and challenges of e-assessment to enhance student learning in higher education. In Cano, E. & Ion, G. (Eds.), *Innovative practices for higher education assessment and measurement* (pp.36-57). Information Science Reference.

Havemann, L., & Sherman, S. (2016). A collaborative approach to improving online assessment and feedback opportunities. In ALT-C 2016: Association for Learning Technology conference: Connect, collaborate, create, 6-8 Sep 2016, Warwick, U.K (Unpublished). [Http://eprints.bbk.ac.uk/16102](http://eprints.bbk.ac.uk/16102) (last accessed 21st Nov, 2017).

Hewson C., & Charlton J.P. (2005). Measuring health beliefs on the Internet: a comparison of paper and Internet administrations of the Multidimensional Health Locus of Control Scale. *Behavior Research Methods, Instruments & Computers*, 37, 691–702. doi: 10.3758/BF03192742.

Hewson, C., Charlton, J., & 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 and Teaching*, 6(1), 37-46. doi: 10.2304/plat.2007.6.1.37.

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), 488-498. doi: 10.1111/j.1365-2729.2011.00473.x.

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education*, 40(1), 1-14.
<http://doi.org/10.1080/02602938.2014.881978> (last accessed 21st Nov, 2017).

JISC (2007). Effective practice with e-assessment: An overview of technologies, policies and practice in further and higher education. HEFCE.

Kingston, N.M. (2009). Comparability of computer- and paper-administered multiple choice tests for K-12 populations: A synthesis. *Applied Measurement in Education*, 22(1), 22-37. doi: 10.1080/08957340802558326.

Lee, C. L., & Huang, M. K. (2014). The influence of computer literacy and computer anxiety on computer self-efficacy: the moderating effect of gender. *Cyberpsychology, Behavior, and Social Networking*, 17(3), 172-180. doi: 10.1089/cyber.2012.0029.

Meyer, A. J., Innes, S. I., Stomski, N. J., & Armson, A. J. (2016). Student performance on practical gross anatomy examinations is not affected by assessment modality. *Anatomical Sciences Education*, 9(2), 111-120. doi: 10.1002/ase.1542.

Mahar, D., Henderson, R., & Deane, F. (1997). The effects of computer anxiety, state anxiety, and computer experience on users' performance of computer based tasks.

Personality and Individual Differences, 22, 683–692. doi: 10.1016/S0191-8869(96)00260-7.

Marriott P. (2009) Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. *British Journal of Educational Technology*, 40, 237–254. doi: 10.1111/j.1467-8535.2008.00924.x.

Osatuyi, B. (2015). Is lurking an anxiety-masking strategy on social media sites? The effects of lurking and computer anxiety on explaining information privacy concern on social media platforms. *Computers in Human Behavior*, 49, 324-332. doi: 10.1016/j.chb.2015.02.062.

Powell, A. L. (2013). Computer anxiety: Comparison of research from the 1990s and 2000s. *Computers in Human Behavior*, 29(6), 2337-2381. doi: 10.1016/j.chb.2013.05.012.

Ricketts C., & Wilks, S.J. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & Evaluation in Higher Education*, 27, 475–479.

Saadé, R. G., & Kira, D. (2009). Computer anxiety in e-learning: The effect of computer self-efficacy. *Journal of Information Technology Education*, 8, 177-191.

Sheader E., Gouldsbrough I. & Grady R. (2006). Staff and student perceptions of computer-assisted assessment for physiology practical classes. *Advances in Physiological Education*, 30, 174–180. doi: 10.1152/advan.00026.2006.

Soffer, T., Kahan, T., & Livne, E. (2017). E-assessment of online academic courses via students' activities and perceptions. *Studies in Educational Evaluation*, 54, 83-93. doi: 10.1016/j.stueduc.2016.10.001.

Stephenson, L.B., & Crete, J. (2010). Studying political behavior: A comparison of Internet and telephone surveys. *International Journal of Public Opinion Research*, 23(1), 24-55. doi: 10.1093/ijpor/edq025.

Stödberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education*, 37, 591–604. DOI:10.1080/02602938.2011.557496. doi: 10.1080/02602938.2011.557496.

van Ballegooijen, W., Riper, H., Cuijpers, P., van Oppen, P., & Smit, J. H. (2016). Validation of online psychometric instruments for common mental health disorders: A systematic review. *BMC psychiatry*, 16(1), 45. doi: 10.1186/s12888-016-0735-7.

Warburton, B. (2009). Quick win or slow burn: modelling UK HE CAA uptake. *Assessment & Evaluation in Higher Education*, 34(3), 257-272. doi: 10.1080/02602930802071080.

Whitelock, D., Gilbert, L., & Gale, V. (2013). E-assessment tales: What types of literature are informing day-to-day practice?. *International Journal of e-Assessment*, 3(1).

Whitworth, D. E., & Wright, K. (2015). Online assessment of learning and engagement in university laboratory practicals. *British Journal of Educational Technology*, 46(6), 1201-1213. doi: 10.1111/bjet.12193.

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological methods*, 18(1), 53. doi: 10.1037/a0031607.

Yang, Z., Chen, H., Zhang, X., Wang, R., & Ding, J. (2016). The online version of the Chinese Intolerance of Uncertainty Scale: Psychometric properties. *Cyberpsychology, Behavior, and Social Networking*, 19(3), 217-222. doi: 10.1089/cyber.2015.0149