The role of attitudinal factors in mathematical on-line assessments: a study of undergraduate STEM students

DOI: 10.1080/02602938.2017.1401976

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Assessment and Evaluation in Higher Education

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Download date: 06. Mar. 2020
The role of attitudinal factors in mathematical on-line assessments: a study of undergraduate STEM students

Elizabeth Acosta-Gonzaga\textsuperscript{a,b} and Niels R.Walet\textsuperscript{a}

\textsuperscript{a}School of Physics and Astronomy, The University of Manchester, Manchester, M13 9PL, UK
\textsuperscript{b}Instituto Politécnico Nacional, Av. Té 950, Granjas México, 08500, Distrito Federal, México

ARTICLE HISTORY
Compiled November 2, 2017

ABSTRACT
This study explores student attitudes to the use of substantive on-line assessments that require mathematical answers. Since there is limited guidance available for their use in a university setting, our goal is to learn what are the important aspects in student acceptance of e-assessments that support learning of mathematical subjects in higher education. To that end we analyse the effects of a variety of attitudinal factors towards such assessments amongst a cross-section of first year students in an English University, using a detailed questionnaire. These students were all previously exposed to on-line assessments containing substantial mathematical work, including testing of and feedback on the algebraic structure of their answers, based on identifiable misconceptions underlying these answers. Since students received highly tailored feedback, the expectation was that the usefulness of this feedback would be the key driver in their usage of educational technology. The results indeed suggest that students find on-line feedback more enjoyable and useful than traditional feedback, but enjoyment and attitude are the two most important factors.

KEYWORDS
online assessments; online formative feedback; technology enhanced learning; mathematics teaching; higher education

Introduction

The continuing expansion of the use of Information and Communication Technology (ICT) gives us many new opportunities, but also poses serious challenges. In the context of learning, a drive towards cost saving and the lure of the new have often led to implementations which are not wholly successful. Many of these implementations have ignored the important attitudinal factors that can affect the usage of learning technology. Information system researchers have studied the importance of such personal factors, e.g., attitudes, beliefs, culture and behaviour, in technology acceptance more than 25 years ago (Davis, Bagozzi, and Warshaw 1992). Such factors have also been investigated in a general setting for the acceptance of educational technology in higher education (Liu, Liao, and Pratt 2009; Teo 2009; Terzis and Economides 2011; Cheung and Vogel 2013). Little work has been done that applies to those areas that require a specialised style of questions and answers, which is especially true for mathematics.

Email: elz.acosta.gonzaga@gmail.com
Email: Niels.Walet@manchester.ac.uk
Most on-line material provides a mixture of what we can roughly call ‘learning materials’ and ‘assessments’. The latter contains both formative (for feedback) and summative (graded) assessments, which are sometimes separate, but one commonly combines summative assessment with formative feedback. The important role of assessment and feedback in learning and teaching in higher education is discussed in detail in the existing literature, e.g., Brown, Bull, and Pendlebury (1997); Gibbs and Simpson (2004); Nicol and Macfarlane-Dick (2006); Bloxham and Boyd (2007); Hooshyar et al. (2016). It is in this process that students’ learning gets consolidated, which in turn produces persistent changes in students’ understanding. This is the reason why educational assessment deserves such attention. The advantages of on-line assessment technologies, either formative or summative, have been summarised by, e.g., Terzis, Moridis, and Economides (2012): “(a) high interaction and adaptation with test-takers, (b) real-time feedback, (c) real-time score reports, (d) more efficient management, setting, and delivering of exams, (e) easier data management, (f) cost reduction, (g) self-evaluation and recognition of students’ strengths and weaknesses”. Also, in order to provide good quality learning, universities regularly make use of innovative technologies to support the teaching, learning and assessment processes. Information technology can clearly be one of those innovation factors, but one should ask oneself what the goal should be. One of the main answers is that it helps in providing access and support to an increasing number of students from in increasingly diverse background. Given that undergraduate classes can consist of several hundred students, it is no longer possible for faculty to meet with individual students and guide their learning, for example by marking out-of-class assignments. This issue, combined with the diversity of students’ academic backgrounds, supports the need for formative electronic assessment (Miller 2009a). Immediate on-line feedback is an essential feature of such assessments: It has positive effects on students’ learning performance; it activates their intrinsic motivation (Dreher, Reiners, and Dreher 2011; Hooshyar et al. 2016; Nikou and Economides 2016) and helps them to achieve their goals (Whitelock and Brasher 2006). Students also have a more positive attitude towards feedback in a computer based assessment context when they receive correct responses immediate and detailed (substantive) feedback (van der Kleij et al. 2012).

Formative electronic assessment provides students with a flexible opportunity to evaluate their understanding of a course, and a mechanism to set their own learning goals and assess their weaknesses and strengths in order to improve their performance in a course (Wilson et al. 2011). Understanding their strengths and weaknesses enables students to focus on their cognitive development. It will also help them to identify key areas of focus for further study. Miller (2009a) shows that a formative e-assessment framework can provide an effective mechanism for providing the required feedback. These studies show that students who use computer-assisted practice quizzes gain significantly higher grades than students who do not. This demonstrates that computer-assisted formative assessments, and specifically online practice tests, can have a positive impact on student performance. “... effective feedback on formative CBAs [Computer Based Assessments] requires that feedback to students be provided immediately after the student has responded to an item.” (Miller 2009b, p. 160). Online formative feedback also serves as an enabler for reflection and self-assessment (Ludvigsen, Krumsvik, and Furnes 2015). It not only improves learning, but also encourages independent learning. Particularly in higher education, formative e-assessments are able to promote self-regulated learning and reflection on what was learned. Students can learn from self-reflection in an individual assessment process and/or from interactions in a peer-assessment (Gipps 2005; Hodgson and Pang 2012). Dreher, Reiners, and Dreher (2011)
remark on the pedagogical benefits obtained, particularly those from feedback, by using e-assessment technologies. The thesis by Jordan (2014) discusses in some interesting detail what makes successful e-assessment for learning, and particularly states that a detailed understanding of the tasks is crucial for engagement.

There are serious difficulties with providing meaningful online assessment and feedback for questions requiring mathematics in their solution and/or bottom-line answer; this is part of the general issue of assessing higher-level cognitive and affective skills (Kuh et al. 2014). There is a substantial literature that discusses the challenging technological and methodological aspects of implementing mathematical assessments, see e.g., Sangwin (2013); Adesina, Stone, and Jones (2015); Haddif and Yerushalmy (2015); Sangwin and Köcher (2016), but little attention has been given to the human factors making such assessments effective. The nature of the field means that there is no reason to believe that the factors playing a role in adopting e-learning in a mathematical context have been fully explored by the more generic studies quoted above. This is especially true in higher education, where we would also like to analyse answers and give feedback on deeper levels of (mis)understanding than in most other settings. This can not easily be achieved by multiple choice questions (Sangwin 2013). The good news is that, even at this level, a sizeable fraction of mistakes in students work can be described algorithmically, and we can thus provide good targeted feedback. We just wish to know if and when this is effective.

There is evidence that using the computer to assess mathematical subjects can bring important advantages; for instance, Dettori, Garuti, and Lenut (2002) have investigated the possibilities of improving mathematical teaching and learning by using technology in various educational contexts. They report that studying mathematical topics supported by technological means leads to a considerable increase in student motivation towards learning mathematics, which also creates a general positive change in their attitude towards the subject (Ursini, Sánchez, and Orendain 2004). There are several studies that investigate this attitude (Galbraith and Haines 1998; Ursini, Sánchez, and Orendain 2004; Reed, Drijvers, and Kirschner 2010), which is associated with determining their willingness to learn. In this work we do not consider this attitude per se. Instead, the focus is on analysing the attitude and intention of using electronic assessment technologies to learn mathematics in the specific context of higher education. There is interesting work by Martínez-Sierra et al. (2016) about the underlying psychosocial representations about mathematical assessment in general, that might map onto e-learning.

Other researchers such as Angus and Watson (2009) illustrate that when students studying mathematics are subject to regular challenging on-line testing, their learning, as measured by an end-of-semester examination, is significantly improved. Indeed, mathematics is particularly suited to an on-line assessment strategy and on-line assessment can provide valuable feedback to students, particularly distance learners (Whitelock and Raw 2003), but there are difficulties extending this to the depth required for university students. At primary school level, Peltenburg, van den Heuvel-Panhuizen, and Doig (2009) show that e-assessment using a dynamic visual tool is able to support students to overcome difficulties in solving subtraction problems. This type of testing allows the teacher to better examine the students’ actions and thinking processes than is possible with a paper-and-pencil test (Whitelock 2009).

Since effective use of on-line feedback is clearly desirable for all stakeholders, it is important to explore the students’ perception of on-line feedback in e-assessment of mathematical material. What we wish to know as educators is: ‘Does targeted formative feedback have a positive impact on the usage of assessment technologies?’; ‘Does
immediate on-line feedback bring more value than traditional feedback?'; ‘Do students find on-line formative feedback useful and/or enjoyable?’.

Our approach is rooted in the belief that on-line assessment done ‘right’ will lead to a positive answer to those questions, but this raises additional ones. It is important to understand what drives and stimulates students to use on-line testing. Therefore we also ask the questions: ‘What are the significant factors that influence usage of assessment technologies for learning mathematics?’; ‘What are the significant factors that stimulate students’ usage of assessment technologies in an advanced mathematical context?’ It is well known that attitudinal factors can play an important role in any technology usage, and these can have long term effects (Rosen and Weil 1995; Venkatesh et al. 2003).

In this paper we report the results of a study carried out with first year undergraduate students at the University of Manchester, UK, who are studying STEM (Science, Technology, Engineering and Mathematics) subjects. We studied the attitudinal factors using an online questionnaire, structured according to a set of empirical and theoretical factors discussed in next section. We construct a model from these factors, which is discussed in the following section. We then discuss the data collection, and analyse the data using our models in the next section. Finally, we draw some conclusions and give some recommendations. In Table 5 we show our full questionnaire, and finally in we illustrate some of the essential features of the type assessments in Fig. A1 in the appendix.

**Theoretical background**

The model developed and applied in this work is empirical, but we use prior studies as a guiding principle in designing it. We wish to understand the influences on a student’s ‘behavioural intention’. The model we use was influenced by the work of Terzis and Economides (2011), who brought ideas of how technology acceptance into the realm of e-assessment. Even though this seems a logical development, there seem to be only a few papers using these ideas (for a recent example see Cigdem and Oncu (2015).

The model is built on the foundations of the technology acceptance model, first proposed by Davis (1989) as a model to predict information systems adoption. The model is based on the earlier Theory of Reasoned Action. The latter was first proposed by Fishbein and Ajzen (1975) as a model to explain and predict the behaviour of people in a specific situation. The theory of reasoned action states that a person’s actual behaviour can be explained by his/her intention and beliefs, and that, intentions can be explained by both his/her ‘attitude’ and ‘subjective norms’. Here ‘attitude’ is defined as “the degree of a person’s favourable or unfavourable evaluation or appraisal of the behaviour in question”. (Fishbein and Ajzen 1975, p. 302) define the ‘subjective norms’ as the “person’s perception that most people who are important to him or her think he or she should or not should perform the behaviour in question”. We have included this as ‘social influence’. See Appendix A for details of the questions that constitute these headings. The effects of ‘subjective norms’ on behavioural intention are direct. The theory of reasoned action also states that ‘attitude’ plays an important role in a person’s ‘behavioural intention’. These two factors were later adopted by the technology acceptance model.

In the technology acceptance model a few new factors that may have an effect on adoption are introduced. The most important ones are ‘perceived usefulness’, which is defined as “the degree to which an individual believes that using a particular system would enhance his/her job performance”, and ‘perceived ease of use’, which is defined as
“the degree to which an individual believes that using a particular system would be free of physical and mental effort” (Davis 1993, p. 477). The technology acceptance model also assumes that ‘perceived usefulness’ itself will be influenced by ‘perceived ease of use’, because if “two systems perform the identical set of functions, a user should find the one that is easier to use more useful” (Davis 1993). Technology acceptance model states that ‘perceived usefulness’ has a direct effect on an individual’s behavioural intention toward using a system, and ‘perceived ease of use’ acts indirectly through its influence on ‘perceived usefulness’ (Davis 1989). That means that ‘perceived usefulness’ mediates the effect of ‘perceived ease of use’ on ‘behavioural intention’. It is important to consider such factors in our work, since it is known that both ‘perceived usefulness’ and ‘perceived ease of use’ can have significant effects on an individual’s behavioural intention to use e-learning systems (Ong, Lai, and Wang 2004; Liu, Liao, and Pratt 2009).

Borrowing from social cognitive theory, a widely accepted and empirically validated theory for understanding and predicting human behaviour and for identifying methods by which behaviour can be changed (Bandura 1986), we also include the construct of ‘self-efficacy’ in our model. In this context, ‘self-efficacy’ is defined as “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances” (Bandura 1986, p. 391). The theory has been applied as a theoretical framework to predict and explain an individual’s behaviour in a variety of contexts involving cognitive, social, motor, health, instructional, and self-regulatory skills. The theory also proposes the importance of ‘performance expectations’ that influence individual behaviour (Bandura 1986). Since performance expectations are defined as “the degree to which a learner believes that using a particular system will help him or her to attain gains in learning performance” (Davis 1993, p. 477), this is similar to ‘perceived usefulness’ in the technology acceptance model, which we have already included. ‘Self-efficacy’ has been shown to be able to predict behavioural change with different types of participants in various settings. Therefore, in order to understand its impact we include it as a factor in our model. In the context of our study it is included as ‘computer-assessment self-efficacy’.

It is well-known that external factors can have an important influence on attitudes; in this context think about, e.g., the accessibility of IT services. We have therefore included the ‘availability of IT’ as a potential factor. It is similar to what Thompson, Higgins, and Howell (1991) and Venkatesh et al. (2003) call “technology and resource facilitating conditions” and Ajzen (1991) mentions as “perceived behavioural control”. In the context of computer use, Thompson, Higgins, and Howell (1991), p. 129, define facilitating conditions as “the provision of support for users may be one type of facilitating condition that can influence system utilization. By training users and assisting them when they encounter difficulties, some of the potential barriers to use are reduced or eliminated”. This is consistent with what (Venkatesh 2000, p. 347) explains “in the context of workplace technology use, specific issues such as the availability of support staff, which is an organizational response to help users overcome barriers and hurdles to technology use, especially in the early stages of learning and use”. In other words, the facilitating conditions include the factors in the environment that shape a person’s perception of ease or difficulty of performing a task (Teo 2012). It embraces factors such as technical support (the provision of help-desks and on-line support services). Technical support has been cited as one of the important factors in the acceptance of technology for teaching and in user satisfaction (Williams 2002; Teo 2012).

By analysing the process of e-assessment, we can identify potentially important factors. One such process is the provision of formative feedback. Whitelock and Brasher
Table 1. Abbreviated labels of the constructs used in our model, together with their full names

<table>
<thead>
<tr>
<th>construct</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUS</td>
<td>Perceived Usefulness</td>
</tr>
<tr>
<td>PEU</td>
<td>Perceived Ease of Use</td>
</tr>
<tr>
<td>CSE</td>
<td>Computer-assessment Self-Efficacy</td>
</tr>
<tr>
<td>SI</td>
<td>Social Influence</td>
</tr>
<tr>
<td>AIT</td>
<td>Availability of IT Services</td>
</tr>
<tr>
<td>EN</td>
<td>Enjoyment</td>
</tr>
<tr>
<td>RF</td>
<td>Received Feedback</td>
</tr>
<tr>
<td>CF</td>
<td>Comparative Feedback</td>
</tr>
<tr>
<td>AT</td>
<td>Attitude</td>
</tr>
<tr>
<td>BI</td>
<td>Behavioural Intention</td>
</tr>
</tbody>
</table>

(2006) state that formative assessment practices using computers encourage immediate feedback, which has positive effects for students in achieving their goals. Also, Terzis (2012), p. 216, points out that “[electronic assessment] provides them immediate feedback to identify their strengths and weaknesses”, see also (Crippen and Brooks 2002; Wilson et al. 2011). Moreover, Dreher, Reiners, and Dreher (2011) mention that the real benefit for students is getting immediate feedback, which enhances their learning performance and also stimulates their intrinsic motivation enjoyment, see Hooshyar et al. (2016). That in turn appears to be linked with playfulness, and is is related to intrinsic motivation (Venkatesh 2000). Terzis (2012), p. 216, argues that the main reasons that learners enjoy using computer-based assessments are: “(1) Learners are able to take the assessment anywhere and any-time using a computer. (2) They are able to take the test as many times as they wish. (3) They feel confident regarding the results’ accuracy and fairness since the computer does not care who the test taker is. (4) They are able to see their results as soon as they complete the assessment. 5) CBA provides them immediate feedback to identify their strengths and weaknesses”. Hooshyar et al. (2016) analysed the effects of online formative assessment on students’ learning, and find that when students who are enjoying and engaging with a learning activity, their attitudes and achievements improve.

We capture these ideas in two more constructs: The first captures the students’ experience with receiving on-line feedback. This is labelled as ‘received feedback’ which stands for “the student’s belief that the support and feedback they receive from the online platform will enhance their learning”. This combines a series of questions about the on-line feedback being precise, clear, helpful and timely. The second construct explores the difference between receiving on-line feedback and face-to-face feedback. We call this ‘comparative feedback’ as a short-hand for “the student’s perception of receiving on-line feedback when it is compared with traditional feedback”.

We want to explore the helpfulness of online formative assessment in attitudes through exploring enjoyment and usefulness. Therefore, we hypothesize that on-line feedback will have a direct positive effect on usefulness and enjoyment (and indirect effects on attitude and behavioural intentions to use web-based testing).

It should thus come as no surprise that we also include an explicit ‘enjoyment’ construct, since enjoyment is known to be a significant reason for interest in using a specific system. Although research on its impact on behavioural aspects comes from an organizational context (Davis 1989, 1993), it seems to be more generally relevant. In an educational setting, enjoyment has been shown to have a positive influence on usage intentions. Venkatesh (2000), p. 348, explains that playfulness is related to intrinsic motivation or the “perceptions of pleasure and satisfaction from performing the behaviour”. This also matches with the work of Davis, Bagozzi, and Warshaw (1992).
who argue that intrinsic motivation refers to the performance of an activity for no apparent reward other than the process of performing the activity *per se*. In his research Venkatesh (2000) explains that playfulness is a construct that is system-independent, supporting this by the fact that users who are more “playful” with computer technologies in general enjoy using a new system just for the sake of using it. Playfulness is expected to be a relevant factor even when the systems perform a rather boring task, since it still involves exploration and discovery. Venkatesh states that, from a theoretical standpoint, a higher level of playfulness will lead to a lower perception of effort. In the context of this research, this could mean that students feel they have to invest less effort when they perceive an electronic educational task (mathematical exercise) to be enjoyable. Since we can consider students as accustomed to using technology in all aspects of their life, it must be relevant to explore the effects of this construct on their attitude and intention to use this for educational purposes. This could thus be a direct determinant to enhance learning of mathematics.

On the other hand, (Davis, Bagozzi, and Warshaw 1992, p. 1113) find that ‘perceived enjoyment’ and ‘perceived usefulness’ mediate the influence of ‘perceived ease of use’ on ‘intention’, explaining that “while usefulness will once again emerge as a major determinant of intentions to use a computer in the workplace, enjoyment will explain significant variance in usage intentions beyond that accounted by usefulness alone.” This clearly requires explanation and investigation. In the context of computer-based assessment, the research of Terzis and Economides (2011) includes perceived playfulness and shows a positive effect on the behavioural intention to use the system. Therefore, we will explore the effects of this construct on attitude and behavioural intentions.

We list a final summary of all our constructs in Table 1, together with a set of abbreviations we shall use in the rest of the paper.

**Research model and hypotheses**

All the constructs proposed in the previous section are included in a hierarchical model, taking into account the network of technology acceptance model. The constructs are summarized in Table 1. From these constructs, using the theoretical and empirical approach from the previous section, we then construct a model as displayed in Figure 1. This model is based on an initial exploration, and contains 14 hypothetical causal relationships between constructs, (H means hypothesis) that will be tested on a set
of data using a recursive Structural Equation Model. This is a widely used multivariate approach, see e.g., the book by Kaplan (2008) for an overview. The method can be used to examine the relationship between theory-based latent variables and their indicator variables by using measurements of directly observable indicator variables (Hair et al. 2010). This assumes that each variable has a linear relation with any antecedent variables, and the recursive nature states that causation is unidirectional. Variables are usually divided into two classes: exogenous, or explanatory variables $x$ and endogeneous or response variables $y$.

A structural equations model assumes a linear form relation between the variables,

$$\sum_{j=1}^{q} B_{ij} y_j + \sum_{j=1}^{p} \gamma_{ij} x_j = \eta_i,$$

where we work in a no-feedback network, where $B$ is a lower-triangular matrix,

$$B_{ij} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ b_{21} & 1 & 0 & \cdots \\ b_{31} & b_{32} & 1 & \cdots \end{pmatrix},$$

and $\eta_i$ is a random fluctuation. We then try to explain as much variance in each construct as possible by the antecedent variables, which is what is called the partial-least squares structural equations modelling, or PLS-SEM.

One of the difficult questions one needs to address in such an approach is its statistical power; unfortunately there is no simple way to estimate this, since the model is highly complex, see for example (Wolf et al. 2013). We have taken an exploratory approach, and have reduced the number of relationships in our model to below the maximum we can fit to our data. In other words, we use the data to partially dictate the model.

Background of collected data

We invited a large number of first year students (about 1000) in the Faculty of Engineering and Physical Sciences at the University of Manchester to answer an anonymous on-line questionnaire. These students are exposed to a variety of on-line materials. Particularly relevant to the subject of this research is the use of mathematical questions using the Stack engine within a Moodle platform (Sangwin 2013) that provides feature-
rich on-line assessments with detailed feedback based on recognising common forms of incorrect reasoning. Specifically to the School of Physics and Astronomy, from which a large fraction of the respondents originate, students also experience online assessment using the “mastering physics” product originally developed by MIT, and now marketed by Pearson. For more details of the way this is used, see Walet and Birch (2012).

Every year, the faculty receives a large number of STEM students. As the incoming students have different mathematical backgrounds, we determine their previous level of mathematical knowledge using a diagnostic test. Practice material for the test, and follow-on tests and remedial material, are all provided within Stack. This allows us to differentiate the material delivered to students and provide additional support when required based on preparation and needs.

Students have the option to practise a large number of online mathematical exercises before answering the diagnostic test. The key characteristic of this material is to provide students with immediate feedback through a web-based platform. If a student answers a question incorrectly, they will be given support based on their mistakes by providing a series of hints to help them obtain the correct solution. Students can practise as long as they want and need to reinforce their mathematical skills. This formative feedback represents a big advantage for on-line testing since it opens the possibility of providing students with customised feedback given that the computer can generate this based on the answer given by the student (López 2009).

While academics may have a wide variety of reasons for selecting particular assessment methods, they need to be aware of their students’ perceptions of these methods and how these influence students’ learning (Iannone and Simpson 2013). However, in order to obtain a complete view of the context, it is also important to consider what students’ think about learning mathematics through technological means. What is their opinion? In order to make the most of what educational technology offers, students should be willing to undertake mathematical exercises and exams using electronic means to make their learning of mathematics more valuable.

All students in STEM subjects at the University of Manchester will have taken mathematics up to final high-school examination level (except in the case of Chemistry, where this is not required). In the English system, that means a mathematics A-level. Each school has its own requirements for entry, and requires different grades in the exams. Nevertheless, all students are required to take the same diagnostic test, and use the same follow-on material. Most students follow mathematics courses in the School of Mathematics, apart from physics students who follow their courses in the School of Physics and Astronomy. In all cases, a variety of on-line learning materials is used. It is important for the Schools to obtain enough data to build a customised learning strategy to provide students with personalised good quality teaching.

Students used the on-line platforms during September and October 2014. At that point we used an on-line survey to gather students’ responses. The survey was completely anonymous and conducted through a personal invitation in a limesurvey server. We obtained 121 full responses, about a 10% response rate, which is on the low side for the expected return for such an online questionnaire. Even though we have no way of knowing who responded, one would like to know how representative this group is. Looking at what information we do have, the gender balance is 26% (31) female 74% (90) male, roughly in line with the gender balance in the faculty. The distribution across fields of study is displayed in Table 2. The majority of students is from the School of Physics and Astronomy, and there we have responses roughly 25% of all students, as we we would expect for a successful online questionnaire. These are weak indicators of a representative sample, and may indicate that the results are slightly exploratory,
Table 2. Distribution of respondents by School of study

<table>
<thead>
<tr>
<th>School</th>
<th>Responses</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Engineering &amp; Analytical Science</td>
<td>20</td>
<td>16.5%</td>
</tr>
<tr>
<td>Chemistry</td>
<td>3</td>
<td>2.48%</td>
</tr>
<tr>
<td>Earth, Atmospheric &amp; Environmental Sciences</td>
<td>2</td>
<td>1.7%</td>
</tr>
<tr>
<td>Electrical and Electronic Engineering</td>
<td>14</td>
<td>11.6%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>6</td>
<td>4.96%</td>
</tr>
<tr>
<td>Mechanical, Aerospace and Civil Engineering</td>
<td>11</td>
<td>9.1%</td>
</tr>
<tr>
<td>Physics &amp; Astronomy</td>
<td>65</td>
<td>53.7%</td>
</tr>
</tbody>
</table>

and would probably benefit from taking a larger sample.

For all 31 questions we invite answers on a five-point Likert scale, where in every case the most negative answer is coded as 1, and the most positive as 5. Using this coding, we show the average and standard deviation for each of the questions in Figure 3 (see Table 5 for the details of the questions). As we can see—this will be studied further below—we have very similar results for the individual questions making up a single construct.

Findings and Results

Applying structural equations modelling uses a two-tier process. First the individual measurements or indicators (in our case answers to individual questions) are combined into constructs or latent variables—the consistency of this process needs to be validated for all latent variables. It measures how well the observed indicators fit the unobserved (latent) variables. In the second stage of the process, we analyse a structural model, which covers the relationships among constructs. The constructs typically represent feelings, attitudes, and opinions of a person. The relationships between constructs are hypothesized in accordance with theoretical and logical reasoning (Götz, Liehr-Gobbers, and Krafft 2010). In this work both the measurements and the structural models were evaluated using SmartPLS 2.0 (Ringle, Wende, and Will 2005).

Measurement Model

We first test the quality of the measurement model. Figure 3 summarises results from the measurement model showing that all indicator loadings are higher than the common threshold criterion of 0.7 for indicator reliability (Hair et al. 2010) (see Table 1 for an explanation of the items). Moreover, cross-loading coefficients were added. Table 3 shows measures of construct and composite reliabilities. We apply some standard tests that are not normally part of the structural equations modelling measurement as well. The values of Cronbach’s alpha, which can be used to analyse the correlation between questions making up a construct, show that all constructs have a good internal reliability except “comparative feedback” and “availability of information technology”. The first of these with a value $\alpha = 0.66$ is only acceptable. The “availability of information technology” has such a small value of $\alpha$ that suggests these questions are not measuring the same variable. On the other hand, the structural equations modelling code evaluates composite reliability and average variance extracted (AVE) as alternative tests of the measurement model. Analysing the values for composite reliability, we find these are all larger than the acceptable cut-off of 0.6 (Bagozzi and Yi 1988). The Average Variance Extracted ranges from 0.54 to 0.81. These values all exceed the rec-
Figure 3. Descriptive statistics for the questions used for each item: The green bar denotes the standard deviation (+/-) on each item; the line in the middle of the green bars denotes the average.

Table 3. Numerical analysis of the structural equations model. The first three columns show construct and convergent validity coefficients for the measurement model (Average Variance Extracted AVE, Composite Reliability CR, and Cronbach \( \alpha \)). The remaining matrix shows the discriminant validity in the off-diagonal entries, which is given by bivariate correlations. On the main diagonal, we show the related square-root of AVE (Average Variance Extracted) of each construct.

| AVE  | CR  | \( \alpha \) | PUS | PEU | CSE | SI | AIT | EN  | RF  | CF  | AT  | BI  |
|------|-----|--------------|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|
| PUS  | 0.70| 0.90         | 0.85| 0.84|
| PEU  | 0.69| 0.87         | 0.78| 0.43| 0.83|
| CSE  | 0.63| 0.87         | 0.80| 0.30| 0.51| 0.79|
| SI   | 0.67| 0.80         | 0.51| 0.40| 0.41| 0.36| 0.82|
| AIT  | 0.54| 0.78         | 0.59| 0.27| 0.51| 0.40| 0.29| 0.73|
| EN   | 0.75| 0.90         | 0.83| 0.66| 0.42| 0.39| 0.39| 0.29| 0.87|
| RF   | 0.68| 0.86         | 0.76| 0.39| 0.31| 0.25| 0.33| 0.38| 0.53| 0.82|
| CF   | 0.60| 0.82         | 0.66| 0.58| 0.35| 0.38| 0.42| 0.28| 0.64| 0.56| 0.77|
| AT   | 0.79| 0.92         | 0.86| 0.67| 0.41| 0.41| 0.37| 0.20| 0.80| 0.53| 0.63| 0.89|
| BI   | 0.81| 0.93         | 0.89| 0.57| 0.34| 0.30| 0.43| 0.15| 0.67| 0.43| 0.46| 0.66| 0.90|

Recommended lower cut-off (Götz, Liehr-Gobbers, and Krafft 2010). Therefore, all values obtained confirm what is called “convergent validity”, the internal consistency of each latent variable.

Table 3 shows values from the measurement model demonstrating that all indicator loadings are higher than the common threshold criterion of 0.7 to reach indicator reliability. We also show the discriminant validity—the fact that the different indicators are statistically independent.

Structural Model

In order to test the statistical significance of the relationships in the model a bootstrap procedure with 200 re-samples was used. In Partial-least squares structural equations modelling, we try to explain as much of the variance in any construct by the variation in antecedent constructs. The results for the model as shown in Figure 2 are summarized in Figure 4. There we give the path coefficient (i.e., the strength of the coupling in the structural equations model), as well as the \( t \)-test values. The relevant critical values for a two-tailed test are: \( t = 1.96 \) corresponds to \( p = 0.05 \), and \( t = 2.58 \) has \( p = 0.01 \). In
this paper we consider $t$-values higher than 1.96 as significant, and reject paths with lower probability than 0.05 and corresponding lower $t$ values below 1.96.

**Discussion**

The overall fit of the model is good with 70% of the total variance explained, which is above the minimum threshold of 60% (Hair et al. 2010). We find that the antecedent constructs explain 53% of the variance in ‘behaviour intention’. At an intermediate stage of the model, 68% of the variance in attitude, 49% of the variance in ‘enjoyment’, 40% of the variance in ‘perceived usefulness’ and 37% of the variance in ‘perceived ease of use’ are explained in a similar way. More importantly, we see that ‘attitude’ and ‘enjoyment’ have an important effect on predicting usage intentions. This shows that, even for detailed mathematical assessments, it is important for students to take pleasure in and appreciate the on-line assessments. It also shows that their opinions and feelings about the assessments make a difference to their usage intentions. On the other hand, we notice that the construct ‘perceived usefulness’ has only an indirect effect through the ‘attitude’. This conclusion matches with the acceptance model for computer based assessment (CBAAM) of Terzis and Economides (2011) who also find that perceived usefulness has no direct effect on behavioural intention to use a computer-based assessment.

‘Attitude’ in turn is predicted by ‘perceived usefulness’ and ‘enjoyment’, but ‘enjoyment’ has a stronger effect than ‘perceived usefulness’. This could indicate that an on-line test has to include fun activities, challenging activities that make students find them enjoyable while they are learning mathematics. We see that creating teaching strategies that include enjoyable activities is essential to enhance student’s learning in mathematics. Making good use of games, quizzes, and other creative approaches to create more enjoyment and interest in learning mathematics is essential.

On the other hand, ‘comparative feedback’ has a strong direct effect on predicting ‘perceived usefulness’ and ‘enjoyment’. This confirms that students value the fact of getting on-line feedback, since they find that it is both useful and enjoyable. This is also demonstrated by looking at the relevance of the indirect effects by evaluating their $t$ values. The effect of ‘comparative feedback’ on attitude and on usage intentions are both highly relevant, as can be seen in Table 4. With respect to enjoyment this finding
matches with the study Moon and Kim (2001) who find that enjoyment has a positive impact on behavioural intentions.

‘Received feedback’ has an important effect on ‘enjoyment’ but not on ‘perceived usefulness’, as can be seen by the direct t-values, $t_{RF \rightarrow EN} = 1.99$ and $t_{RF \rightarrow PUS} = 0.56$. This seems to indicate that students enjoy receiving on-line feedback in mathematical assessment. The indirect effects on attitude ($t_{RF \rightarrow AT} = 1.85$) and usage intentions ($t_{RF \rightarrow BI} = 1.64$) demonstrate that this factor is not crucial to trigger a favourable behavioural intention, see Table 4 for more details.

Perceived ease of use has a strong influence on perceived usefulness ($t_{PEU \rightarrow PUS} = 2.93$) and enjoyment ($t_{PEU \rightarrow EN} = 2.52$). This also contributes to the strong indirect effect on attitude ($t_{PEU \rightarrow AT} = 3.00$) and usage intentions ($t_{PEU \rightarrow BI} = 2.90$). These match what we mentioned earlier: perceived enjoyment and perceived usefulness mediate the effects of perceived ease of use on intention. This means that this factor is essential to obtain positive attitude and usage intentions. This is proven by following the chain (PEU → EN → AT → BI) that could indicate that if students perceive an assessment as easy to use, they are more likely to have an enjoyable experience, and they are more willing to use it. The link (PEU → PUS → AT → BI) shows that when assessment technology is easier to operate, it is more useful, and therefore, students are more willing to use it.

Availability of information technology and computer-assessment self-efficacy have a strong effect on perceived ease of use. This could indicate that students who encounter some difficulties during an on-line test (regarding system’s operation or questions’ content) need technical support such as help-desks, on-line support services, guidance by the IT staff and faculty to overcome these situations. Therefore, when technical assistance is provided, it is more likely to find using the on-line environment easier. This is supported by Terzis and Economides (2011) who explain that in the context of computer-based assessment, the availability of information technology determines perceived ease of use. We also see that students who feel comfortable using computers, will find it easier to do mathematical on-line assessments.

These results show that the indirect relationships AIT → AT, AIT → BI, AIT → EN, AIT → PUS, CF → AT, CF → BI, CSE → AT, CSE → BI, CSE → EN, CSE → PUS, PEU → AT, PEU → BI are all significant for the model. Furthermore, using the measurement model, especially the significance of indicator variables, we see that items AT1, AT2, BI2, BI3, CF3, EN3, and PUS1 are the most important in the model. If we quickly look at what these specific questions ask, we see that AT1 “I like doing on-line test and exams in subjects that require mathematical answers” and AT2 “I look forward to those aspects of my learning of mathematics that require me to use on-line assessment” suggest that we need to make sure that the current tests satisfies the students expectations as well as possible. The questions CF3 “on-line feedback helps me better understand mathematical subjects”, EN3 “doing mathematical on-line tests is enjoyable” and PUS1 “I find on-line tests useful to support my learning of
mathematical subjects” are also statements about the quality of the current tests, and clearly show those aspects test development should focus on—enjoyment, usefulness, and quality feedback. The importance of items BI2 “I would like to continue my use of on-line assessment to support my learning of mathematics” and BI3 “all things considered, I expect to continue doing on-line test or exams to assist my learning of mathematics” are more troublesome—we do know it is a problem, and as yet largely unsolved, to develop online assessments for material beyond that used in the first year of the English university curriculum.

Conclusions

In this study we have explored the factors playing a role in adopting formative online assessments in an advanced mathematical context. To this end, we examined student attitudes on the use of high-feedback assessments that require mathematical answers. Our aim was to learn what are the important aspects in a design of more effective e-assessments that support learning of mathematical subjects in a higher education setting, and to see whether just providing high quality feedback makes an effective assessment.

We have found that feedback indeed has a strong effect on predicting the perceived usefulness and enjoyment. This demonstrates the most important message from this study: we need to combine utility with enjoyment in a successful assessment, even in “dry” subject such as applications of advanced mathematics. This is also demonstrated by looking at the strong positive indirect effects of what we term ‘comparative feedback’ on students’ attitude and usage intentions. We note that the attitude to received feedback has a strong influence on enjoyment but not on perceived usefulness. This indicates the rather surprising result that students find the experience of receiving feedback in an on-line (mathematical) test more enjoyable than useful. Also, the indirect effects on attitude and behavioural intention have a small $t$ value, and as such the process of receiving feedback is less important to the attitude and acceptance of on-line testing than we expected. However, we find that enjoyment is a key factor to effective design of online mathematical assessments. This is consistent with what (Davis, Bagozzi, and Warshaw 1992, p. 1113) point out “enjoyment will explain significant variance in usage intentions beyond that accounted for by usefulness alone.” Thus we see that even though the nature of mathematical e-assessment is special, the acceptance seems to roughly follow the pattern we see for general e-assessment (Terzis and Economides 2011).

Nevertheless, our results also reveal that attitude and enjoyment are the most important factors influencing students’ usage intention. To our surprise usefulness does not have a direct effect—in a future study we hope to report on instructors attitudes, but indications are that usefulness for students is an important driver for instructors. On the other hand, usefulness has an indirect effect: together with enjoyment it strongly predicts attitude.

We conclude that an effective educational design of online assessments, even in mathematics, has to include activities, exercises, questions and feedback that students find enjoyable to motivate them to effectively use assessment technologies. Providing good feedback is part of this, but there is much more to enjoyment!

In its turn, ‘perceived ease of use’ is an important factor influencing ‘usefulness’ and ‘enjoyment’. This also contributes to its strong indirect effect on ‘attitude’ and ‘usage intentions’. Therefore, ‘enjoyment’ and ‘perceived ease of use’ are the most powerful
factors for predicting usage intentions. We know that ‘availability of information technology’ and ‘computer-assessment self-efficacy’ have a strong effect on perceived ease of use. Taking into account these factors can be the best way to design a mathematics e-assessment activity for UK students.

We see that even though our model is based on the technology acceptance model, the model we end up with is more complex, and includes a number of affective factors which have proven to be important. This is not surprising since we know that such factors can be quite relevant, and we should really look at a multitude of factors, especially those that make a task enjoyable.

The role of enjoyment suggests the question whether doing mathematics online in this way is perceived as having a game-like element for students. It is well known that gamification of assessment activities can have a very positive effect on students’ outlook (Kapp 2012), and an intriguing study by Attali and Arieli-Attali (2015) analyses how points for tests can be interpreted as a game in the context of mathematics assessment. This is an interesting question for future study.

Anyone who designs an assessment activity must make sure that it is enjoyable. A good example is the main engine used with our physics students, mastering physics by Pearson. Doing an online search will lead quickly to websites where students voice their distaste of this very capable tool (Mastering critique 2014). We would therefore recommend to start by giving adequate training to the students before using the assessment engine, so that they can make most effective use of the assessment. One needs to check carefully under which circumstances, both hardware and software based, such assessments have problems, and steer students to a sufficient number of IT resources that deal with these assessments well. Assessments need to be well designed, designing tasks that students find more enjoyable than useful. It would probably be a good idea to study what tasks students do find enjoyable, and which they dislike. Feedback needs to be good and consistent, but is not as important as we thought at the start of this study. For advice on designing effective feedback, the work by Jordan (2014) is a source we would recommend the reader to consult.

Clearly our work has some limitations: the questionnaire shown in Table 5 covers only a limited number of points of interest, and we may want to extend it in the light of the results of the study, especially to elucidate the role of feedback and enjoyment. Repeats of the study with a similar but larger cohort of students would be of interest. Since we have not completely defined the elements of enjoyment in an assessment, some qualitative studies using structured interviews and observations should definitely be considered. We would also like to study the instructors’ attitudes, since we believe that an effective assessment process must have balanced expectations at both ends. Finally, it might be interesting to repeat this study in a different cultural context, to see whether our conclusions remain valid.

Acknowledgements

The first author would like to thank The National Polytechnic Institute (IPN) and The National Council for Science and Technology of Mexico (CONACYT) for support for this research.
<table>
<thead>
<tr>
<th>Table 5. A list of all the questions in the questionnaire answered by students, grouped under their respective categories.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived Usefulness (PUS)</strong></td>
</tr>
<tr>
<td>PUS1</td>
</tr>
<tr>
<td>PUS2</td>
</tr>
<tr>
<td>PUS3</td>
</tr>
<tr>
<td>PUS4</td>
</tr>
<tr>
<td><strong>Perceived Ease of Use (PEU)</strong></td>
</tr>
<tr>
<td>PEU1</td>
</tr>
<tr>
<td>PEU2</td>
</tr>
<tr>
<td>PEU3</td>
</tr>
<tr>
<td><strong>Computer-assessment Self-Efficacy (CSE)</strong></td>
</tr>
<tr>
<td>CSE1</td>
</tr>
<tr>
<td>CSE2</td>
</tr>
<tr>
<td>CSE3</td>
</tr>
<tr>
<td>CSE4</td>
</tr>
<tr>
<td><strong>Social Influence (SI)</strong></td>
</tr>
<tr>
<td>SI1</td>
</tr>
<tr>
<td>SI2</td>
</tr>
<tr>
<td><strong>Availability IT Services (AIT)</strong></td>
</tr>
<tr>
<td>AIT1</td>
</tr>
<tr>
<td>AIT2</td>
</tr>
<tr>
<td>AIT3</td>
</tr>
<tr>
<td><strong>Enjoyment (EN)</strong></td>
</tr>
<tr>
<td>EN1</td>
</tr>
<tr>
<td>EN2</td>
</tr>
<tr>
<td>EN3</td>
</tr>
<tr>
<td><strong>Received Feedback (RF)</strong></td>
</tr>
<tr>
<td>RF1</td>
</tr>
<tr>
<td>RF2</td>
</tr>
<tr>
<td>RF3</td>
</tr>
<tr>
<td><strong>Comparative Feedback (CF)</strong></td>
</tr>
<tr>
<td>CF1</td>
</tr>
<tr>
<td>CF2</td>
</tr>
<tr>
<td>CF3</td>
</tr>
<tr>
<td><strong>Attitude (AT)</strong></td>
</tr>
<tr>
<td>AT1</td>
</tr>
<tr>
<td>AT2</td>
</tr>
<tr>
<td>AT3</td>
</tr>
<tr>
<td><strong>Behavioural Intention (BI)</strong></td>
</tr>
<tr>
<td>BI1</td>
</tr>
<tr>
<td>BI2</td>
</tr>
<tr>
<td>BI3</td>
</tr>
</tbody>
</table>
References


chronous e-learning systems in high-tech companies.” *Information and Management* 41: 795–804.


Appendix A. Example of the style of questions considered

As an example of the capabilities we expose our students, consider the simplified problem discussed in Figure A1. There we show a highly simplified example of the type of feedback given to students, some of which is based on known common misconceptions and mistakes (the term malrules, which stands for “a consistent application of incorrect rules” is often used), such as differentiating when asked to integrate, mistakes in carrying through factors in functions (i.e., functions of functions), etc. In mastering physics our approach is much the same; in addition we also use analysis based on the presence of terms containing the relevant physical parameters.

We assume that the student is asked to perform the indefinite integrate the function \( \cos(3x) \), and have shown some of the feedback that can be given. Note that this is actually an extremely simple example of the type of feedback a student may be exposed to.

Figure A1. Flow diagram representation of the feedback that can be given in a simplified stack question.

Does answer differentiate correctly?

Is the answer the derivative?

Does the answer contain integration constant?

Feedback:

Does answer differentiate correctly?

Is the answer the derivative?

Is the answer off by a factor 3?

Feedback:

Feedback:

Feedback:

Feedback:

Feedback:

Feedback: Well done