Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates

How to cite:
Nguyen, Quan; Rienties, Bart; Toetenel, Lisette; Ferguson, Rebecca and Whitelock, Denise (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. Computers in Human Behavior, 76 pp. 703–714.

For guidance on citations see FAQs.

© 2017 Elsevier Ltd

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.1016/j.chb.2017.03.028

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.
Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates

Quan Nguyen, Bart Rienties, Lisette Toetenel, Rebecca Ferguson, Denise Whitelock

Institute of Educational Technology, The Open University, UK
Bank Julius Baer & Co, Zurich, Switzerland

ABSTRACT

Many researchers who study the impact of computer-based assessment (CBA) focus on the affordances or complexities of CBA approaches in comparison to traditional assessment methods. This study examines how CBA approaches were configured within and between modules, and the impact of assessment design on students’ engagement, satisfaction, and pass rates. The analysis was conducted using a combination of longitudinal visualisations, correlational analysis, and fixed-effect models on 74 undergraduate modules and their 72,377 students. Our findings indicate that educators designed very different assessment strategies, which significantly influenced student engagement as measured by time spent in the virtual learning environment (VLE). Weekly analyses indicated that assessment activities were balanced with other learning activities, which suggests that educators tended to aim for a consistent workload when designing assessment strategies. Since most of the assessments were computer-based, students spent more time on the VLE during assessment weeks. By controlling for heterogeneity within and between modules, learning design could explain up to 69% of the variability in students’ time spent on the VLE. Furthermore, assessment activities were significantly related to pass rates, but no clear relation with satisfaction was found. Our findings highlight the importance of CBA and learning design to how students learn online.

1. Introduction

The field of Learning Design has emerged during this century. It seeks to develop a conceptual and descriptive framework of teaching practices, making use of technology, in order to make learning designs explicit, sharable, and reusable (Conole, 2012; Kirschner, Strijbos, Kreijns, & Beers, 2004). Since assessment is a common element of learning design, it is of great importance to understand how assessment activities are configured, and to represent them in an explicit, sharable, and reusable manner. In terms of capturing the design of assessment activities, an increasing number of educators use computer-based assessment (CBA). CBA can be defined as assessment presented using digital means and submitted electronically. CBA offers many advantages over traditional forms of assessment as it is electronically built, and therefore generates user data on assessment activities.

CBA allows teachers to compare the designs of their assessment approaches across different tasks and modules. For example, Toetenel and Rienties (2016) compared 157 modules at The Open University (OU) and found that, on average, 21.50% of students’ total workload was allocated for assessment, although substantial variation (SD = 14.58%, range 0%—78%) was found amongst these modules. By representing CBA in an explicit way, educators can learn from each other by comparing, reusing, and adapting peers’ CBA designs in their own learning context.

The impact of CBA designs on the learning processes of students may be better understood with the support of learning analytics, which make use of the digital traces of learners’ interactions in a virtual learning environment (VLE) that are preserved in log-files. Recent learning analytics research has found that the way in which teachers design tasks and assessments at a micro level (within one assessment or task: see for example Greiff, Wüstenberg, and Avvisati (2015)) and a macro level (across various assessments within or across modules: see for example

Keywords:
- Computer-based assessment
- Learning design
- Learning analytics
- Academic retention
- Learner satisfaction
- Virtual learning environment
Lockyer, Heathcote, and Dawson (2013)) can influence how students are engaging with CBA tasks and their academic performance (Koedinger, Booth, & Klahr, 2013; Rienties & Toetenel, 2016; Toetenel & Rienties, 2016).

For example, in a fine-grained log file study of one Programme for International Assessment (PISA) CBA task on climate control that compared complex process solving amongst 16,219 children in 44 countries, Greiff et al. (2015) found substantial differences in individual strategies. These strategies had a significant impact on individual performance and also highlighted differences in problem-solving strategies between countries. In a follow-up study of 1476 Finnish children completing nine complex problems, Greiff, Niepen, Scherer, and Martin (2016) found that there was an optimal level of effort spent on these tasks and consecutive performance, as well as a negative relation between the frequency of changes made in learning strategy and performance. These micro-level CBA studies provide rich and complex understandings of how students engage with specific assessment tasks (Koedinger et al., 2013; Vandewaetere, Desmet, & Clarebout, 2011), and may help teachers to make decisions about intervention both on individual tasks and on a broader, module-based level.

Similarly, but at a more macro-level of analysis, Rienties and Toetenel (2016) studied the activity of 111,256 students on 151 modules at the OU using multiple regression models. The study found that learning designs strongly predicted VLE behaviour and performance of students. In follow-up research, Rogaten, Rienties, and White洛克 (2016) compared the learning gains of 17,700 students on 110 Science and Social Science modules and found individual differences in learning gains as well as significant differences in assessment practices within and across OU modules. Module characteristics accounted for 6%–33% of variance in students' initial achievements and 19%–26% of variance in subsequent learning gains. In other words, recent preliminary findings suggest that how teachers design CBA influences how students learn over time on both a micro and a macro level.

By aligning the designs of CBA with fine-grained data relating to students' engagements with the VLE, together with their satisfaction and performance, educators may be equipped with valuable insights as to how their students are “reacting” to the design of CBA (Koedinger et al., 2013; Rienties & Toetenel, 2016; Toetenel & Rienties, 2016). While, to the best of our knowledge, the study by Rienties and Toetenel (2016) was the first to link aggregate learning design data of CBA and other activities on a large scale with actual student behaviour and cognition across a large number of modules, the study did not deal with how educators use CBA on a week-by-week basis. Aligning how educators balance CBA on a weekly basis with what students actually do in the VLE can not only advance our insights into CBA, but may also help to strengthen the links between learning analytics and CBA.

Therefore, firstly we aim to understand how educators design and implement CBA on a weekly basis. Specifically, we will investigate how educators have allocated time for seven types of learning activity (assessment, assimilative, finding information, communication, productive, interactive, and experiential) at both module level (for 74 modules) and weekly level (for 37 modules).

Secondly, by building on previous research (Rienties & Toetenel, 2016; Toetenel & Rienties, 2016), we will investigate the impact of CBA designs on students' behaviours in the VLE using log-files data, satisfaction levels, and module pass rates. This element of the study is of particular interest as a large body of assessment literature has indicated that assessment drives learning (Bearman et al., 2016; Segers, Dochy, & Cascallar, 2003), but limited empirical evidence on a macro level is available to confirm this claim. Our analysis will use a combination of visualisation techniques and fixed effect models on 72,377 registered students studying 74 modules across different disciplines.

Finally, while learning analytics may highlight key correlations and even causation in relation to learning design, CBA, student engagement and learning, it is important to be aware that learning always takes place in a context. Therefore, our third and final aim is to unpack how six teachers designed their modules and understand how their weekly design decisions influenced student learning, using a case-study approach.

2. Computer-based assessment, learning design and learning analytics

CBA has a lot of potential to demonstrate how students learn and solve complex tasks (Greiff et al., 2015; Ras, Whitelock, & Kalz, 2015; Tempelaar, Rienties, & Giesbers, 2015). It offers many advantages when compared to more traditional forms of assessment. In distance-learning settings the most relevant benefits are: speed, cost-reduction (Terzis & Economides, 2011), automatic feedback provision (Ras et al., 2015), record keeping, and more authentic interactive assessment options. These options include intelligent tutoring (Koedinger et al., 2013), adaptive CBA (Tempelaar et al., 2015; Vandewaetere et al., 2011), and authentic virtual labs (Scherer, Meßinger-Koppelt, & Tiemann, 2014).

In a study of six online and two blended courses, Agudo-Peregrina, Iglesias-Pradas, Conde-González, and Hernández-García (2014) found that interactions with CBA tools, interactions with peers and teachers, as well as active participation were significant predictors of academic performance on online courses. Similarly, in a blended course in mathematics and statistics that relied intensively on CBA, Tempelaar et al. (2015) found that behaviour and mastery in CBA assessments were good predictors of academic performance. In contrast, most metrics not linked to CBA that can easily be gathered from a VLE (such as time spent and number of downloads) provided limited insight into student progression.

In a recent study of computer programming courses using CBA design, Brito and de Sá-Soares (2014) found that a high frequency of CBA, at weekly level, was one of the most effective ways of setting students on the route to success in an introductory computer programming course. Weekly CBA may help to speed up the cycle of productive failure (Kapur, 2008) — “fail fast to learn sooner” (Brito & de Sá-Soares, 2014) — because it can provide automated feedback. In a follow-up fine-grained study of 1080 students using worked examples (i.e., accessing the help function of a particular task) during CBA in mathematics, Tempelaar et al., 2017 found that students with sub-optimal learning strategies tended to use these worked examples at the end of their learning process, while students with effective metacognitive strategies used these at the beginning of their learning process. In other words, there is some support for the idea that CBA in conjunction with learning analytics data enables researchers and educators to unpack and predict successful learning.

Most studies that have conceptualised and tested different variations of CBA influence on student engagement and learning outcomes have used a single module (Tempelaar, Rienties, & Nguyen, 2017 ; Tempelaar et al., 2015) or a single unit of assessment context (Greiff et al., 2015, 2016). With the advent of learning analytics (Ferguson & Buckingham Shum, 2012; Rienties & Toetenel, 2016) and learning design (Conole, 2012; Rienties & Toetenel, 2016), researchers can now compare how teachers
across a range of modules and disciplines are designing their modules. This could allow us to test which learning designs lead to better student engagement and learning.

2.1. CBA as a part of learning design

Conole (2012, p. 121) described learning design as “a methodology for enabling teachers/designers to make more informed decisions in how they go about designing learning activities and interventions, which is pedagogically informed and makes effective use of appropriate resources and technologies”. Learning design focuses on what students do as part of their learning, rather than on the content that is delivered by the teaching. Within this journal, it is increasingly evident that learning design is an important driver for learning (Hernández-Leo, Moreno, Chacon, & Blat, 2014; Moreno-Ger, Burgos, Martinez-Ortiz, Sierra, & Fernández-Manjón, 2008; Rienties & Toetenel, 2016).

For example, in a large scale study of 111,256 students on 151 modules at the OU using multiple regression models, Rienties and Toetenel (2016) found that learning designs strongly predicted VLE behaviour and performance of students. Additionally, Li, Marsh, Rienties, and Whitelock (2016) carried out a study of 422 undergraduate blended and online courses that involved logistical regression modelling of learner satisfaction using 232 variables including learning design, learner characteristics and assessment types. They found that learning design decisions significantly influenced learner satisfaction. While the most important key driver for learner satisfaction was the quality of teaching materials, the second most important driver was the quality of assessment (Q36: Overall, I was satisfied with the assessment on this module).

In addition, assignment instructions (Q9: The instructions on how to complete the assignments were clear) and assignment completion (Q10: Taking part in optional exercises or activities to test my understanding helped me to learn) were consistently among the top 10 drivers of learner satisfaction (Li et al., 2016).

These studies linked large datasets across a range of modules in both online and blended learning settings (Arbaugh, 2014; Li et al., 2016; Rienties & Toetenel, 2016). Together, they demonstrated that it is possible to obtain important and generalisable insights that extend beyond the research findings associated with a single module or discipline. This can be done by carrying out a cross-sectional study that analyses the impact of learning design on assessment strategies, learner satisfaction and academic performance across a range of modules.

2.2. Research questions

While substantial advances in our understanding of assessment strategies have recently been made (Bearman et al., 2016; Ras et al., 2015; Tempelaar et al., 2015), few studies have examined effective assessment strategies across a number of modules. Furthermore, in a recent study Rienties and Toetenel (2016) found that although learning design is widely studied as a way to improve course design (Armellini & Ayiegbay, 2010; Koedinger et al., 2013; MacLean & Scott, 2011), few institutions have captured and updated these data in order to reflect on how these courses are delivered to students.

Research question 1 focuses on how teachers design and use CBA in online distance learning, and how these designs relate to the broader notions of learning design developed by Conole (2012). By addressing research question 2 we aim to understand how CBA decisions made by teachers influence student engagement in the VLE, and how this engagement may influence student satisfaction and pass rates. In comparison to Rienties and Toetenel (2016), a unique new contribution of this study is that we will investigate the impact of weekly CBA design decisions in 37 modules on student engagement in research question 2. This may be conceptually and empirically important, as CBA tasks in, say, week 4 may have an impact not only on student engagement in week 4, but also on engagement before and after the assessment activity. Therefore, using regression and fixed effects models in the first part of this study we will focus on two research questions and build on the work of Rienties and Toetenel (2016) by applying a learning analytics perspective to a large dataset.

1) To what extent do educators use individual assessment designs and how do these relate to learning design principles?
2) To what extent are educators’ assessment design decisions associated with student engagement in the VLE, student satisfaction and pass rates?

Two recent meta-analyses on learning analytics (Papamitsiou & Economides, 2014, 2016) and a large-scale review of learning analytics impact studies for the European Commission (Ferguson et al., 2016) indicated that relying extensively on large data and on quantitative proxies of learning can make it difficult to understand the context in which learning and CBA strategies take place. Therefore, research question 3 prompts a more fine-grained analysis of the assessment and learning design strategies of six modules designed for first-year undergraduates.

3) Which, if any, weekly assessment strategies are associated with student engagement over time?

3. Method

3.1. Setting and participants

This study took place at the OU, a distance-learning institution with an open-entry policy and the largest university in the UK. As previous research has found substantial differences between postgraduate and undergraduate learning designs (Li et al., 2016; Rienties & Toetenel, 2016), this study included only undergraduate modules that have run since 2014 with at least 200 registered students per module. In total, 72,377 students were enrolled in these 74 undergraduate module presentations, with an average of 978 (range: 207–3707) registered students. These modules related to various disciplines (25% in Science & Technology, 22% in Arts & Social Sciences, 14% in Business & Law, 9% in Education & Languages, and 30% in other disciplines).

There were more female students (54%) than male students (46%) studying these 74 modules. The majority of these students were from the UK (96%) and declared their ethnicity to be ‘white’ (87%). Students varied considerably in age, with 25% under 25 years old, 36% aged 26–35, 21% aged 36–45, 12% aged 45–55, and 6% aged 56 and over. More than half of them were working full-time (52%), while 19% were working part-time, 8% were looking after the home/family, and 6% were unemployed and looking for a job. Regarding learners’ qualifications, there are no formal academic entry requirements at undergraduate level at the OU. In this study, 41% of the students had A levels or equivalent (suggesting they had two or more years of post-compulsory schooling), 31% had less than A levels (suggesting they had not progressed beyond compulsory schooling), 22% had higher education degrees, and 4% had a post-graduate qualification. On average, 9% of the students had a reported disability.

Longitudinal data relating to learning design were available on a weekly basis for 37 of the 74 module presentations (these modules had 45,190 registered students). While learning designs had originally been modelled for each two- to eight-week block of study, the
research findings of (Rienties & Toetenel, 2016; Toetenel & Rienties, 2016) prompted the learning design team to model learning designs on a week-by-week basis.

3.2. Instruments

3.2.1. Learning design mapping

A university-wide learning initiative to use learning design data was developed by Conole (2012) for quality enhancement purposes. This initiative introduced a process of “module mapping” or “coding learning activities”. The taxonomy used to do this has been enhanced by the Jisc-sponsored Open University Learning Design Initiative (OULDI) (Cross, Galley, Brasher, & Weller, 2012), which took place over five years in consultation with eight higher education institutions. This initiative involved analysing and providing visualisations of the learning activities and resources associated with each of the university’s modules (Rienties & Toetenel, 2016).

For a detailed description of the mapping process and reliability of this approach, we refer to our previous article in this journal (Rienties & Toetenel, 2016).

In contrast to instrumental design, learning design is process based (Conole, 2012); following a collaborative design approach in which practitioners make informed design decisions with a pedagogical focus. To do this, they use representations in order to build a shared vision, which is intended to ensure that these practitioners jointly provide a cohesive learning experience for students.

As illustrated in Table 1, Assessment activities include all work that is focused on summative, formative or self-assessment. According to the assessment handbook of The Open University UK (2016), common types of assessment include: assignments, oral or practical assessments, projects, examinations, dissertations, and portfolios. Each module usually includes two assessment components. The first is known as continuous assessment because it takes place throughout the module. Continuous assessment includes tutor-marked assignments (TMAs), and computer-marked assignments (CMAs). CMAs may be submitted via online forms or may be interactive (iCMAs). End-of-module assessments (EMAs) are the second assessment component. These involve examinations (usually hand-written), or other pieces of work such as dissertations, projects, or portfolios. As this takes place in a distance-education context, the assessment is mediated by technology and so is considered to be CBA. This definition could be problematic because in some cases learning and assessment materials are available both online and offline, or learners can print the assessment materials and submit by post. However, they meet our definition of CBA because all the assessment materials are provided through digital means and the majority of learners submit their work online.

The Assimilative activities in Table 1 relate to tasks in which learners attend to discipline-specific information. These tasks include reading text (online or offline), watching videos and listening to audio files. The next five types of activity can be used to create an interactive environment in which learning with others can take place (Ferguson & Buckingham Shum, 2012). By Finding and handling information provided by others, rather than using only content provided by educators, learners are able to take responsibility for their own learning. During Communicative activities, students communicate with others about module content (Arbaugh, 2014; Kirschner et al., 2004), and during Productive activities they build and co-construct new artefacts. Experiential activities provide learners with opportunities to apply their learning in a real-life setting. Finally, interactive activities also provide opportunities to apply learning, but in safe settings such as simulations (Scherer et al., 2014).

The seven types of learning activity were measured in terms of the duration in hours that was allocated for each type of activity. This was determined through a detailed and comprehensive the mapping process between learning designers and instructors (see Rienties & Toetenel, 2016 for further details). These measurements were captured at a module level (for 74 modules) and at a weekly level (for 37 modules). The number of credits to be gained determined the total workload of each module, which is the sum of the time allocated for all seven types of learning activity. Generally speaking, each credit is associated with 10 h of study (so 30 credits = 300 h and 60 credits = 600 h). However, the actual workload can be different, and depends on each module's learning design. Of the 74 modules covered by this study, two offered 15 credits, 40 offered 30 credits, and 32 modules offered 60 credits.

3.2.2. VLE data: student engagement

In line with Tempelaar et al. (2015), two different types of data were gathered for each module from the university’s Moodle VLE: average time spent on the VLE per week (in minutes); and average time spent per session on the VLE (in minutes). Subsequent derivations of these two types of data per week were recorded from week -3 until week 40 (data streams typically start three weeks before the official start of the module). More fine-grained tracking data were available for some modules, providing information about types of content, materials and ICT tools (such as wikis, videoconference and discussion forums). However, given the diversity in usage and because not all modules used all the ICT tools, we focused on aggregate user statistics per week across the VLE as our proxy for student engagement.

Table 1

<table>
<thead>
<tr>
<th>Learning design activities.</th>
<th>Type of activity</th>
<th>Example</th>
<th>Mapped/Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>All forms of assessment (summative, formative and self assessment)</td>
<td>Write, Present, Report, Demonstrate, Critique.</td>
<td>Based upon estimated average time spent per week, unless specified by educator.</td>
</tr>
<tr>
<td>Assimilative</td>
<td>Attending to information</td>
<td>Read, Watch, Listen, Think about, Access.</td>
<td>Based upon word count or time per audio/video clip.</td>
</tr>
<tr>
<td>Finding and handling information</td>
<td>Searching for and processing information</td>
<td>List, Analyse, Collate, Plot, Find, Discover, Access, Use, Gather.</td>
<td>Indicated by educator or estimated by Learning Design team.</td>
</tr>
<tr>
<td>Communication</td>
<td>Discussing module-related content with at least one other person (student or tutor)</td>
<td>Communicate, Debate, Discuss, Argue, Share, Report, Collaborate, Present, Describe, Create, Build, Make, Design, Construct, Contribute, Complete.</td>
<td></td>
</tr>
<tr>
<td>Productive</td>
<td>Actively constructing an artefact</td>
<td>Practice, Apply, Mimic, Experience, Explore, Investigate.</td>
<td></td>
</tr>
<tr>
<td>Experiential</td>
<td>Applying learning in a real-world setting</td>
<td>Explore, Experiment, Trial, Improve, Model, Simulate.</td>
<td></td>
</tr>
<tr>
<td>Interactive/adaptive</td>
<td>Applying learning in a simulated setting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from (Rienties & Toetenel, 2016)
3.2.3. Learner satisfaction
Since its foundation nearly 50 years ago, the OU has consistently collected learner feedback in order to improve its learning experience and learning designs. The Student Experience on a Module (SEAM) survey is employed as part of this process, just as with other learner satisfaction instruments (Onwuegbuzie et al., 2007; Zerihun, Beishuizen, & Os, 2012). This standard questionnaire is sent to all students who are still registered at the end of the module.

Following our analysis of key drivers of the learning experience of 115,000 students (Li et al., 2016), for this analysis we used the aggregate scores of five core items (out of 40) from the SEAM survey that have been shown to drive learner satisfaction. These five items measured students’ satisfaction with regard to (1) teaching materials, (2) assessment on module studied, (3) advice and guidance provided for module study, (4) integration of materials, and (5) career relevance, scaling from one to five in which one means “definitely agree” and five means “definitely disagree”. On average, 81% of the respondents were satisfied with their learning experience (SD = 11%), with a minimum of 39% and a maximum of 97% (Table 2). The average response rate for 74 modules was 31.67% (SD = 7%, min = 18.71%, max = 53.30%).

3.2.4. Pass rate
On average, there were 978 registered students per each module (SD = 824), with a minimum of 205 and a maximum of 3707. The pass rate was calculated as the percentage of registered students who completed and passed the module (Range: 34%–90%; Mean = 63.58, SD = 10.22).

3.3. Data analysis
All data were collected at an aggregate, module level and were anonymised in line with the university’s ethical guidelines. Based on module ID and the year and month in which the module was presented (some modules run twice a year), the learning design data for 74 module presentations were merged with the available VLE and pass rate data, which allowed us to analyse the combined data for 71 modules.

In preparation for the panel analysis of 37 modules (those with weekly longitudinal data), a Hausman test was used to differentiate between a fixed effects and a random effects model. This test checks whether the coefficients estimated by the random effects estimator are the same as the ones estimated by the consistent fixed effects estimator (Hausman, 1978). Our result supported the assumption of correlation between observation errors and predictors. For this reason, a fixed effects model was used as it removes the effect of time-invariant characteristics to assess the net effect of the predictors on the outcome.

Variance inflation factor (VIF) was computed after each model to check for multicollinearity. All VIFs for the predictors were smaller than 2.00, indicating there was no significant correlation among the independent variables. In other words, there was little overlap of measurements among seven types of learning activity. We opted to report unstandardized coefficients because all the explanatory variables were measured in the same unit (hours). Thus, it was more informative to report the original metrics. Our analysis and visualisations were performed using Stata 13 and Tableau 10.1 respectively.

In line with recent recommendations (Creswell & Plano-Clark, 2011), we employed a mixed-method research design. We applied fixed effect models to a wide range of undergraduate modules at the OU, followed by a triangulated case study approach using six selected first-year modules. Yin (2009) emphasised that a case study investigates a phenomenon in depth and in its natural context. The purpose of a case study is to gain in-depth information about what is happening, why it is happening, and the effects of what is happening.

Table 2
Descriptive statistics of key variables of 74 modules.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assimilative (in hours)</td>
<td>74</td>
<td>167.14</td>
<td>60.73</td>
<td>6.00</td>
<td>299.90</td>
</tr>
<tr>
<td>Information</td>
<td>74</td>
<td>8.34</td>
<td>12.42</td>
<td>0.00</td>
<td>88.50</td>
</tr>
<tr>
<td>Communication</td>
<td>74</td>
<td>8.01</td>
<td>9.65</td>
<td>0.00</td>
<td>43.83</td>
</tr>
<tr>
<td>Productive</td>
<td>74</td>
<td>33.47</td>
<td>28.73</td>
<td>0.00</td>
<td>131.60</td>
</tr>
<tr>
<td>Experiential</td>
<td>74</td>
<td>8.19</td>
<td>21.78</td>
<td>0.00</td>
<td>164.00</td>
</tr>
<tr>
<td>Interactive</td>
<td>74</td>
<td>4.20</td>
<td>8.77</td>
<td>0.00</td>
<td>51.15</td>
</tr>
<tr>
<td>Assessment</td>
<td>74</td>
<td>55.48</td>
<td>30.67</td>
<td>0.00</td>
<td>174.25</td>
</tr>
<tr>
<td>Total workload</td>
<td>74</td>
<td>224.83</td>
<td>95.43</td>
<td>90.60</td>
<td>515.30</td>
</tr>
<tr>
<td>Average time per week (in minutes)</td>
<td>74</td>
<td>90.85</td>
<td>54.58</td>
<td>0.00</td>
<td>304.50</td>
</tr>
<tr>
<td>Average time per visit</td>
<td>74</td>
<td>18.33</td>
<td>7.46</td>
<td>0.00</td>
<td>36.93</td>
</tr>
<tr>
<td>Registration (no. of students)</td>
<td>74</td>
<td>978.07</td>
<td>824.20</td>
<td>205.00</td>
<td>3707.00</td>
</tr>
<tr>
<td>Pass rate (%)</td>
<td>74</td>
<td>63.58</td>
<td>10.22</td>
<td>34.47</td>
<td>90.64</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>74</td>
<td>81.22</td>
<td>10.98</td>
<td>38.96</td>
<td>97.34</td>
</tr>
</tbody>
</table>

Note: data aggregated on module level.

Fig. 1. Scatterplot of the 74 modules showing percentage pass rate of registered students who started the module in relation to the percentage of study time allocated for assessment activities. Note: The six case study modules are marked. R-Squared = 0.097.
Using visual inspection of Fig. 1 and Fig. 2, we selected six first-year modules (three Arts/Social Science and three Science/Mathematics) as case studies. Each of the six represented a group of modules within the scatterplots in terms of intensity of assessment and relative performance in terms of pass rates. A detailed review of the relevant learning design data for these modules (balance of activities, assessment types, assessment spread and overall workload) was used to enhance the quantitative analysis of these learning designs.

4. Results

4.1. Assessment design and learning design principles

Research question 1 requires exploration of the extent to which educators used individual CBA designs and how these related to the broader learning design principles of Conole (2012). Descriptive statistics (Table 2) indicated that on average, students were allocated 55.48 h (SD = 30.67) for assessment activities in each module. This accounted for 25.20% of the total workload, with a range of 17.00–174.25 h (Table 2). On the 37 modules for which weekly learning design data were available, students’ total workload was on average 7.80 h per week, of which 1.90 h were allocated for assessment, 3.90 h for assimilative activities, 0.20 h for finding information, 0.23 h for communication, 1.30 h for productive activities, 0.08 h for experiential activities, and 0.17 h for interactive activities.

To further unpack the complexity of CBA design on a week-by-week basis, we visualized the seven types of learning activity in 37 modules over 34 weeks (Fig. 3). At a glance, these modules included relatively few assessment activities (blue) in the first three weeks, whereas more assessment activities were assigned at the end of the module, in particular from week 29 onwards where end-of-module assessment was the dominant learning activity. Peaks in assessment can be seen in week 5, in week 10 (before Christmas), in week 23 (before Easter), and at the end of each module.

Bearing in mind how assessment activities had been designed, we went on to examine how CBA designs related to other learning activities. In Table 3, correlation analyses on a weekly level indicate that assessment activities were significantly and negatively correlated with assimilative, communication, information, productive, and experiential activities. To examine the effects of other learning activities on assessment activities, a fixed effect model with assessment as the dependent variable was conducted (Table 4). Assimilative ($B = -0.70$, SE = $0.02$, $p < 0.01$), information ($B = -0.68$, SE = $0.09$, $p < 0.01$), communication ($B = -0.73$, SE = $0.10$, $p < 0.01$), productive ($B = -0.39$, SE = $0.05$, $p < 0.01$), and experiential ($B = -0.36$, SE = $0.15$, $p < 0.05$) activities were negatively associated with assessment activities. In other words, using a week-by-week perspective, a rise in the amount of time allocated for assessment activities was significantly linked to a fall in the time allocated for all other types of activity, except for interactive. These negative relationships imply that educators introduced assessment activities while reduced other activities in order to avoid an overwhelming workload. This is an important finding as previous research using only aggregate data showed limited relations between assessment and other learning activities (Rienties & Toetenel, 2016).

4.2. Impact of assessment design on student engagement, student satisfaction and pass rates

4.2.1. Students’ engagement in VLE

After exploring the CBA designs, we started to investigate how assessment design is associated with students’ activity in the VLE (RQ2). Fixed effect models were conducted with the average time spent on VLE per visit and per week as dependent variables. For each predictor, four models were applied. First, we ran a normal OLS regression model. Second, we used the fixed effect model to control for the unobserved heterogeneity of time. Third, we controlled for the fixed effect between modules. Finally, we controlled for the fixed effects of both time and modules. Since assimilative activities account for most of the workload, they were set as the baseline. Therefore, the following results should be interpreted relative to assimilative activities.

Table 5 shows that assessment activities were positively and significantly related to the average time spent in the VLE per week in all four models. In Models 1 and 2, the effect of assessment activities was almost the same ($B = -4.98$, SE = $0.57$, $p < 0.01$ and $B = -5.09$, SE = $0.59$, $p < 0.01$ respectively). The effect of assessment activities became smaller in Model 3 and Model 4 when differences between modules were taken into account. On average, an
additional hour allocated for assessment activities was associated with 2.47 (SE = 0.47, p < 0.01) and 2.80 (SE = 0.47, p < 0.01) minutes increase in the average time spent on the VLE per visit in Models 3 and 4 respectively. Other strong predictors of student engagement in the VLE were communication in all four models and interactive activities in Models 1 and 2. Overall, learning design activities explained up to 58% of the variability in student engagement in the VLE per week when controlling for the heterogeneity within and between modules. Additionally, communication activities were positively associated with time on VLE per visit in all models, while productive, experiential, and interactive activities had a significant effect in Models 1 and 2 only. Overall, by taking into account the heterogeneity within and between modules, learning design was able to explain 69% of the variability in time spent on the VLE per visit (Model 4).

### 4.2.2. Pass rates and students' satisfaction
Having identified the effects of CBA designs on VLE engagement, we continued to investigate how assessment activities influence pass rates and students' satisfaction. Scatterplots were created that plotted assessment activities against pass rates (Fig. 1) and against learning satisfaction (Fig. 2). The figures revealed a diversity of approaches to assessment. On average, 25% of students' workload was allocated for assessment activities. However, this average represents a broad range. Some modules, such as Case Study 6 (CS6) and CS4 (see section 4.3 for detailed description of these modules), had a relatively low reliance on assessment activities. Some, like CS5 (which allocated 57% of the total workload for assessment), had a strong focus on assessment activities. Other modules, like CS1 and CS3, were nested around the average of 25%, while CS2 was one standard deviation above the mean. Follow-up linear regressions (Table 7) indicated that CBA assessment activities were positively related to pass rates, while there was no significant relation between assessment activities and satisfaction. Student satisfaction was primarily predicted negatively by student-active learning activities, namely experiential, productive, finding information, and communication. In line with Rienties and Toetenel (2016), student satisfaction was positively associated with pass rates when controlling for the differences between modules.

### Table 3
Correlation matrix of learning design and VLE engagement at weekly level.

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Assimilative</td>
<td>-0.46**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Communication</td>
<td>-0.12**</td>
<td>0.17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Information</td>
<td>-0.12**</td>
<td>0.08**</td>
<td>0.17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Productive</td>
<td>-0.29**</td>
<td>0.16**</td>
<td>0.13**</td>
<td>0.17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Experiential</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Interactive</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. VLE per week</td>
<td>0.20**</td>
<td>0.01</td>
<td>0.27**</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.16**</td>
<td></td>
</tr>
<tr>
<td>9. VLE per visit</td>
<td>0.12**</td>
<td>0.10**</td>
<td>0.22**</td>
<td>0.04</td>
<td>0.07**</td>
<td>0.04</td>
<td>0.09**</td>
<td>0.84**</td>
</tr>
</tbody>
</table>

N = 37 modules (1088 data points).

*p < 0.05, **p < 0.01.

### Table 4
Fixed effect model of assessment and other learning activities.

<table>
<thead>
<tr>
<th>DV</th>
<th>(1) Assessment</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assimilative</td>
<td>-0.70** (0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>-0.68** (0.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>-0.73** (0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productive</td>
<td>-0.39** (0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiential</td>
<td>-0.36 (0.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive</td>
<td>0.14 (0.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.50** (0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations 1088
Adjusted R-squared 0.59

Unstandardized coefficients
Standard errors in parentheses.

*p < 0.05, **p < 0.01.

### Table 5
Fixed effect model of VLE engagement per week predicted by learning design activities.

| DV – VLE per week | Unstandardized coefficients | | | | | |
|-------------------|-----------------------------|---|---|---|---|---|---|
| Models | (1) OLS | (2) FE_week | (3) FE_module | (4) FE_module_week |
| Assessment | 4.98** (0.57) | 5.09** (0.59) | 2.47** (0.47) | 2.80** (0.47) |
| Information | 2.40 (2.64) | 3.23 (2.60) | -0.72 (1.98) | 0.15 (1.94) |
| Communication | 26.29** (2.66) | 26.29** (2.62) | 16.54** (2.16) | 17.44** (2.11) |
| Productive | 1.75 (1.14) | 1.73 (1.12) | -1.84 (1.04) | -1.83 (1.03) |
| Experiential | 3.57 (3.83) | 4.49 (3.78) | -2.07 (2.98) | -0.99 (2.91) |
| Interactive | 11.57** (2.23) | 11.25** (2.20) | -0.33 (1.81) | -0.46 (1.78) |
| Constant | 95.66** (2.91) | 95.30** (2.85) | 110.6** (2.46) | 172.1** (10.50) |

Observations 1088
Adjusted R-squared 0.58

Standard errors in parentheses.

*p < 0.05, **p < 0.01.

Baseline: assimilative.
satisfaction was positively related to assimilative learning activities (i.e., the benchmark variable).

4.3. Assessment strategies in relation with student engagement over time

After carrying out the quantitative analyses, we selected six case study modules based on their positions in Figs. 1 and 2. Additional fine-grained qualitative (e.g., assessment brief for the assessment task) learning design data were available for these modules. For each of these case study modules, the balance of activities was considered to see how the time spent on assessment compared to the balance of other activities. Findings are summarised in the assessment strategy column in Table 8.

As Table 8 outlines, there were substantial variations in the main assessment methods across the six case studies (i.e., assignment, face-to-face exam, final assignment and interactive computer-marked assessment). However, assignments were the main form of assessment in each of the six case studies. In most cases, students were asked to answer a number of questions and then write a short essay. Two of the six case studies (CS4, CS5) included interactive CBAs, which tested students’ factual knowledge. These modules varied widely. Furthermore, CS4 included practice quizzes to help students prepare for their interactive CBAs, whilst CS3 included assessment preparation tasks for each of the assignments. The only module in which assessment took place almost continuously was CS4, which performed well in terms of satisfaction but not in terms of student pass rates. However, when looking at the learning design data in more detail, each case study module included 6–28 assessment activities. This is an interesting finding, as many educators would expect to see at least one assessment activity per week in courses of this length.

Figs. 4–6 illustrate the learning design of the three case studies over 34 weeks. At a glance, assessment and assimilative activities were the dominant designs. As was evident in Fig. 3, there were gaps in weeks 12–13 and week 24 due to Christmas and Easter breaks, as well as breaks in the schedule to prepare for the next learning activities. Assessments were typically conducted every three weeks throughout the course, except in the case of CS4 (Fig. 6). For CS4 a continuous line of assessment can be seen (the blue block) because students were working towards four assessments. In line with our fixed effect analyses, teachers seemed to include assessment at the expense of other learning activities.

A positive relation between assessment and time spent on VLE per week is reflected in Figs. 4–6, as VLE engagement usually went up in weeks with assessment activities. The workload fluctuated considerably between assessment weeks and non-assessment weeks. For example, in CS3 (Fig. 5) the workload in the assessment weeks was similar to that in non-assessment weeks. However, in CS2 (Fig. 4) students were expected to study for more time during assessment weeks than in non-assessment weeks. In other words, both the qualitative narrative, the time dedicated to assessment activities, and actual student behaviour highlight substantial differences in the way these six modules supported learners and provided assessment for and of learning.

### 5. Discussion

Assessment and feedback are key drivers for learning (Conole, 2012; Scherer et al., 2014; Segers et al., 2003). With the omnipresence of CBA in blended, distance learning and MOOCs, one would expect that some consistent patterns of design of assessment across modules on a macro level could be identified. Pedagogy and learning design are of key importance in online learning (Conole, 2012; Kirschner et al., 2004) but limited research has made explicit links between learning design, learning behaviour, and
the references to colour in this article.

Fig. 4. Longitudinal visualisation of learning design (coloured blocks) and average students’ engagement (red line) in the VLE each week for CS2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Longitudinal visualisation of learning design (coloured blocks) and average students’ engagement (red line) in the VLE each week for CS3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. Longitudinal visualisation of learning design (coloured blocks) and average students’ engagement (red line) in the VLE each week for CS4.

learner performance due to a lack of integrated, linked empirical data (Kirschner, Sweller, & Clark, 2006; Rienties & Toetenel, 2016; Toetenel & Rienties, 2016). In the UK, the OU has been leading developments in data gathering through a comprehensive learning design system, which is used by specialists in order to code learning activities. This means that time spent on a range of learning activities can be quantified and analysed in an empirical manner.

Our first, and perhaps most interesting, finding is that although time allocated for assessment varied between 13% and 57% of overall study time, CBA seemed to be balanced with other learning design activities, as our fixed effect model (Table 4) indicated negative betas for all six learning design activities. In other words, in line with good assessment practice (Boud & Falchikov, 2006; Koedinger et al., 2013; Segers et al., 2003) on a weekly level educators did seem to balance assessment with the other six learning design activities, rather than using it to increase student workload in those weeks.

Our second important finding was that the design of weekly
learning design activities significantly predicted VLE engagement by students, explaining up to 69% of variance (research question 2). While assessment did significantly predict students’ online behaviour over time, communication and interactive activities also helped to explain variations in VLE engagement within and across modules. Furthermore, the relative frequency of time allocated for assessment activities was significantly related to student pass rates, but no clear relation with satisfaction was found, in line with previous research (Rientes & Toetenel, 2016). Recent research has highlighted a need for learning analytics researchers to take time into consideration (Greiff et al., 2016; Kovanovic et al., 2015; Tempelaar et al., 2015).

Our findings indicate that VLE engagement is not only substantially influenced by learning design, but is particularly influenced by how educators within and across modules balance their learning design activities on a week-by-week basis. Building on conceptual work related to learning design (Armellini & Aiyegbayo, 2010; Conole, 2012; Lockyer et al., 2013; Toetenel & Rientes, 2016), this study provides further empirical evidence that, in order to understand how students are learning online and how to analyse fluctuations in their online behaviour over time, learning design should play a central role in LA modelling in general, and CBA modelling in particular (Rienties & Economides, 2014).

Finally, our six case studies highlight the complexities that educators face when developing appropriate learning designs (research question 3). Very different sets of learning activities were built into the designs of these six modules, and these designs significantly influenced how students engaged with the VLE (see Figs. 4–6). Two case studies (CS4 and CS5) incorporated assessment most throughout their module, in line with the learning outcomes. This is considered to be good practice: teaching and assessment should be aligned. However, these modules did not perform better than the modules in which assessment activities were related to more defined assessment points in the module.

Our fine-grained analyses of these six case studies showed that the modules had very different assessment strategies. This was a surprise, given that extensive production and validation processes are in place at the OU. When students are confronted with different assessment strategies with each new module they take, this requires them to adapt their learning strategies (Bearman et al., 2016; Boud & Falchikov, 2006). While some variation in learning design and particularly in assessment may encourage interest and motivation for some students, others with less flexible learning strategies may struggle to cope if a previously successful learning strategy is not effective for the next assessment (Tempelaar et al., 2017; Terzis & Economides, 2011; Wolff, Zdrahal, Nikolov, & Pantucek, 2013).

6. Limitations and future work

One obvious limitation of our research is that we did not include more micro level fine-grained analyses of actual student behaviour on CBA tasks, such as those done by Greiff et al. (2016). Given the size of its student population, the OU is currently only able to keep fine-grained data for the modules on which predictive analytics engines are being piloted instead of individual student data (Wolff et al., 2013). In the near future, we aim to analyse the types of strategy that individual students employ when dealing with different CBA designs. This will help to link our macro-understanding of learning design and assessment with fine-grained individual user behaviour.

A second limitation is that we used rather crude measurements of course success (pass rates and satisfaction). Whether students were actually learning and to what extent good pass rates and high satisfaction scores are actually an accurate reflection of appropriate learning designs and CBA can be debated. The diversity of disciplines and students on Level 1 modules (Rientes & Toetenel, 2016) can influence how satisfaction is interpreted and measured, which undermines the comparability of course success across modules. As a consequence of lacking individual student data, we were not able to perform a multiple-group confirmatory factor analysis across modules to address the issue of measurement invariance of the latent construct, satisfaction. Multi-level models taking into consideration both learner characteristics (Tempelaar et al., 2015; Terzis & Economides, 2011) and learning designs (Hernández-Leo et al., 2014; Lockyer et al., 2013; Toetenel & Rientes, 2016) will need to be explored in the future in order to determine which CBA designs are appropriate for which type of learner and discipline.

Further longitudinal research following students over different modules across a qualification is needed in order to understand how flexible and effective learners are in adjusting their learning strategies to the assessment strategy of a particular module, and whether there is a need to harmonise assessment designs across modules. It is clear from the fine-grained learning design analysis undertaken that further research into the use of formative assessment and ways of coding this information would be helpful in order to understand the role of assessment spread, load and forms of CBA and their impact on student outcomes and satisfaction.

7. Conclusion

This study examined the design of CBA and its effects on student engagement in the VLE, satisfaction, and pass rates using a combination of visualisations, and mixed methods on 74 first-level undergraduate modules at the OU. The first finding indicated that, on average, assessment activities accounted for 25% of the total workload, with great variability across modules. Moreover, the workload on other activities decreased when assessment activities were introduced. This implied that educators aimed to balance the total workload when designing CBA.

Secondly, assessment activities were associated with more time spent in the VLE, since the majority of assessment activities required computer use. Learning design in general could explain up to 69% of the variance in VLE engagement. Modules with higher relative frequency of assessment activities are associated with higher pass rates, while no clear relation was found with regard to satisfaction.

Finally, the six case study modules illustrated the diversity of assessment strategies across disciplines in terms of assessment types and approaches. Further visualisations support our findings above as educators reduced the workload on other activities when introducing assessment, and the time spent in VLE increased considerably in assessment weeks.

Our study contributes to CBA innovation by approaching CBA from a learning design perspective using learning analytics. While CBA has been studied extensively in terms of micro-analysis on individual assessment items (Greiff et al., 2015, 2016) and assessment strategies within a module (Brito & de Sá-Soares, 2014; Tempelaar et al., 2015; Terzis & Economides, 2011), the connection between CBAs and learning designs across a large number of distance learning modules has not been explicitly examined in the past. Therefore, by investigating CBA and other learning activities in tandem, we provided a broader picture of how teachers design assessment activities within their modules and across time.

In line with recent reviews of learning analytics (Ferguson et al., 2016; Papanitsiou & Economides, 2014, 2016), we encourage researchers to look beyond “cold” learning analytics data (such as weekly learning design activities, CBAs and student engagement), as a rich diversity of practice seemed to be present when
contrasting six case studies. While our findings highlight a strong relation between weekly learning design activities and student engagement, it is important to bring teachers and educational researchers on board to unpack the complexities of learning.

In terms of practical implications, assessment and feedback are high on the priority list for students and educators, as these link directly to student success and to the success of a course, programme, faculty and university. Some policy makers have already made moves intended to improve the effectiveness of teaching (Ferguson et al., 2016). For example, a Teaching Excellence Framework has been introduced in the UK, and it is likely that measures related to assessment will be used as key indicators. In order to explain how satisfaction and assessment activities are linked and which elements of assessment (balance of activities, spread through module material or assessment methods) have a significant impact on student outcomes, we need to combine research data and institutional data and work together in order to solve this complex puzzle.

Acknowledgements

We would like to thank the reviewers for their detailed and excellent feedback. In particular, we are grateful for the continuous encouragement from the special editor of this special issue on Innovations in CBA, Prof Samuel Greiff Quan Nguyen is supported and funded by the Leverhulme Trust Research Project Grant “Open World Learning”.

References


