ORIGINAL RESEARCH ARTICLE

The E-Design Assessment Tool: an evidence-informed approach towards a consistent terminology for quantifying online distance learning activities

Helen Walmsley-Smitha*, Lynn Machinb and Geoff Waltonc

aAcademic Development Unit, Staffordshire University, Stoke-on-Trent, Staffordshire, UK; 
bSchool of Life Sciences and Education, Staffordshire University, Stoke-on-Trent, Staffordshire, UK 
cDepartment of Languages, Information & Communications, MMU, Manchester UK

(Received: 10 July 2018; Revised: 31 October 2018; Accepted: 18 December 2018; Published: 8 February 2019)

Online distance learning (ODL) continues to expand rapidly, despite persistent concerns that student experience is poorer and retention lower than for face-to-face courses. Various factors affect ODL quality, but the impact of recommended learning activities, such as student interaction activities and those involving feedback, have proven difficult to assess because of challenges in definition and measurement. Although learning design frameworks and learning analytics have been used to evaluate learning designs, their use is hampered by this lack of an agreed terminology. This study addresses these challenges by initially identifying key ODL activities that are associated with higher quality learning designs. The learning activity terminology was tested using independent raters, who categorised the learning activities in four ODL courses as ‘interaction’, ‘feedback’ or ‘other’, with inter-rater reliability near or above recommended levels. Whilst challenges remain for consistent categorisation, the analysis suggests that increased clarity in the learning activity will aid categorisation. As a result of this analysis, the E-Design Assessment Tool (eDAT) has been developed to incorporate this key terminology and enable improved quantification of learning designs. This can be used with learning analytics, particularly retention and attainment data, thus providing an effective feedback loop on the learning design.

Keywords: learning activity; technology enhanced learning; terminology; online learning; learning design

Introduction

Higher education students are increasingly combining face-to-face learning with online distance and blended courses. In the United States 6 million students took at least one online course as part of their degree, which represented 30% of students in 2015 (Allen and Seaman 2017). In the UK, 10% of students in 2012–2013 were distance learners (Garrett 2015). The University of Edinburgh intends to include at least one fully online course in every undergraduate programme by 2025 (Haywood 2016), a trend that is likely to continue because of demand for more flexible learning.

Online distance learning has its critics. High retention rates are often used as a measure of overall course quality (Lenert and Janes 2017), but retention is of concern,
often being much lower than the equivalent face-to-face version (Simpson 2013). For example, the UK Open University retention rate was 22% in 2010 despite its specialism in distance learning (Simpson 2010). A range of possible factors affecting retention have been examined, ranging from learner-specific factors including age, gender, prior educational experience, levels of motivation and self-efficacy to institutional and course-specific factors including support available, course structure and the development of a learning community (e.g. Bawa 2016). A study of distance learning course designs identified that some courses did not contain quality course features, for example, synchronous activities or projects (Lenert and Janes 2017). Furthermore, ‘[w]hat is missing is the trajectory that would complete the feedback loop: the built-in evaluation of designs to see whether they achieved the expected outcomes’ (Mor, Ferguson, and Wasson 2015, p. 224). A feedback loop would enable exploration of the specific impact online learning designs have on students’ learning and make possible recommendations for effective learning activities to enhance learning and retention.

Evaluation of learning designs is hampered by a lack of shared vocabularies for pedagogic practice (Currier et al. 2006, section 2.2, no pagination). To achieve effective evaluation through a feedback loop requires ‘a more widely used language or framework for sharing Learning Designs’ (Dalziel et al. 2016, p. 260). For Laurillard (2012) it is an educational imperative to describe and represent online learning designs so that they can provide feedback to tutors about their effectiveness.

Research objectives
A variety of common educational terminology is used by tutors to describe learning activities, but the extent to which they agree with the meaning and application is not known. This study therefore aimed to provide a reliable quantitative framework for categorising online activities by means of the following:

Objective 1: identifying types of effective online learning activities that support retention
Objective 2: testing terminology used to describe learning activities to identify the extent to which different users agree
Objective 3: developing the e-Design Assessment Tool (eDAT) utilising this terminology to describe and quantify learning activities

Literature review: effective online learning activities
Levels of feedback and interaction in the course are two course design features often cited as having a significant impact on retention and each are discussed in the following sections.

Interaction
Support for interaction in learning comes from social constructivist learning theory (Vygotsky and Cole 1978). Moreover, Croxton’s (2014) meta-analysis indicates that both level and quality of interaction influence online retention.

The literature includes different ways to define and measure interaction (Wanstreet 2006). ‘Transactional distance’ (Moore and Kearsley 2011) suggests physical and psychological distance between tutor and student is the main difficulty
of distance learning. Moore (1989) identified three types of interaction: student–student, student–tutor and student–content. A fourth type of student–interface interaction has been proposed (Hillman, Willis, and Gunawardena 1994). Despite the wide use of Moore’s interaction types, there is no clear agreement on how to measure them (Ekwunife-Orakwue and Teng 2014). The following examples demonstrate how different surveys and data have been used to explore the impact of interaction on student retention.

The Community of Inquiry model (Garrison 2011) for online learning emphasises interaction between students and tutors, referred to as ‘social presence’. Liu, Gomez, and Yen (2009) used the Social Presence and Privacy Questionnaire to measure social presence and identified it as a significant predictor of course retention and final grade. ‘Resonance’ was used as a way to increase social presence by the use of video lectures, and analysis of the video access data suggested that this increased retention (Geri 2012).

An analysis suggested that the number of communication activities designed into a course was the primary predictor for retention (Rienties and Toetenel 2016). They examined 151 ODL courses and calculated the time students were expected to spend on ‘communication’ using Conole’s learning activity taxonomy (Fill and Conole 2005).

A combination of data mining of forum posts and the use of their own student survey showed a positive correlation between student satisfaction and interaction rates (Fasse, Humbert, and Rappold 2009). However, the challenge of isolating individual features of online courses to assess the impact of retention was highlighted by Godwin, Thorpe, and Richardson (2008). They found no significant difference between courses with a variety of interaction patterns when comparing retention and attainment.

Ekwunife-Orakwue and Teng (2014) found a positive correlation between tutor–student interaction and retention by using student satisfaction and computer self-efficacy surveys. Hawkins et al. (2013), using their own survey, found that feedback, procedural interaction and social interaction positively impacted on course completion.

A web-based peer-tutoring system called Online Peer-Assisted Learning, which enhanced interaction by supporting students tutoring each other, also resulted in improved retention (Evans and Moore 2013). The study used social network analysis and the Student Assessment of Learning Gains survey. The use of web-conferencing and structured group tasks achieved high retention as measured by course data and a course experience survey (Thorpe 2008). Interaction in collaborative group assignments using synchronous and asynchronous discussion as well as social media activities increased retention, according to data in the virtual learning environment (VLE) student activity log (Fisher and Baird 2005). Furthermore, frequency, rather than degree, of student interaction was identified as a positive marker for retention when VLE data was analysed (Shelton, Hung, and Lowenthal 2017).

Few studies have explored the impact of student–content interaction in online learning, making this an area for possible further development (Xiao 2017). The use of the eDAT as discussed in the following will enable further research in this area.

**Feedback**

Assessment and feedback activities are common in online learning. There are a variety of types, including formative individual and group tasks, online quizzes and tests, simulations, provision of model answers and summative assignments. Hattie’s (2003)
meta-analysis of teacher effectiveness found that giving students feedback was identified as a highly effective intervention.

The impact of regular feedback to student postings was highlighted by Stott’s (2016) case study, suggesting that low levels of student engagement and satisfaction may be the result of a lack of tutor feedback. A series of analytical writing assignments with feedback increased retention on a PhD programme by 39% (Sutton 2014). A cross-unit diagnostic that gave feedback to online learners from different learning units also had a positive effect on retention (Lin et al. 2014).

Bonk and Khoo (2014) highlighted the negative impact on online retention when prompt and individual feedback was not given. Choi et al.’s (2013) survey identified that a lack of feedback from tutors was a key reason for students not re-enrolling.

A systematic review of the impact of peer-assessment in online learning indicated that this ‘improves performance of students in learning environments in over 60% of the evaluated articles’ (Tenório et al. 2016, p. 103). A course redesigned to include regular tests with automatic feedback increased attainment and reduced withdrawal (Sancho-Vinuesa, Escudero-Viladoms, and Masia 2013).

Interaction and feedback are inherently linked: a tutor giving feedback to students is a form of interaction, and interactions with students provide feedback to tutors on how students are progressing (Hatzipanagos and Warburton 2009).

Representing learning designs

The impact of course design features on retention can be investigated using the Learning Design Conceptual Framework (Dalziel et al. 2013). Dalziel argues that Learning Design can be used in fine-grained comparisons in educational research and that there is a need ‘to keep trying to develop a broadly accepted representational framework(s)’ (Dalziel et al. 2016, p. 256). Laurillard agrees:

Perhaps the attempt is doomed. But without it there is no basis for the comparative analysis of the range of conventional and digital teaching methods that will tell us how they may best be used to support student learning. That is an imperative for our education systems now, so we have to try.

(Laurillard 2012, Chapter 5, no pagination)

Learning design representations are ways to represent or ‘codify’ learning designs to help online tutors and learning designers analyse and innovate, facilitate software developers to instantiate lessons in software or share designs with others (Conole 2013). Representations can include practice-based, conceptual, abstract or technical learning designs and those based on a specific theoretical approach. They can represent individual lessons or whole courses and provide different lenses to explore specific features including the nature of the task, the tools, resources or pedagogic principles. The most common type of representation is textual; other examples include content and course maps, pedagogy profiles, task swim lanes (visualisations) and learning outcome maps (Conole 2013). However, each representation uses different terminology and formats, some embedding pedagogic guidance and others not. The learning design representations in Table 1 illustrate the variety of terminology used to describe learning activities by different tools.

This variety of learning activity terminology is challenging for learning designers when evaluating the effectiveness of learning designs. For example, the Open
University mapping project used Conole’s taxonomy (Cross et al. 2012) to create a learning activity map over many courses. However, the authors commented on the difficulty of applying these terms, saying the process was ‘subjective’ and that they held ‘regular meetings to improve consistency’ (Rienties, Toetenel, and Bryan 2015, p. 316). Swan edited and applied six of Reeves’ (1996) 14 pedagogical dimensions to her work describing MOOC pedagogies; she also commented that raters needed a number of discussions to agree on their application (Swan et al. 2015). Similarly, Laurillard observed that although tutors were able to map their own activities to a taxonomy, they were unable to agree when asked to map another tutor’s task (Charlton, Magoulas, and Laurillard 2012). Analysis of a number of US online courses used a rubric for raters to score each of four key elements on a three-point scale; it experienced similar difficulties (Jaggars and Xu 2016). Even very simple terms seem to cause difficulties; for example, some users thought there was ambiguity between ‘resource’ and ‘support’ in the AUTC representation (Agostinho 2011).

A group of learning designers conducted an interesting study to apply different learning design tools to a single lesson plan to ‘represent’ the design. Their challenges and varied results highlight the lack of consistency in learning design tools (Persico et al. 2013). This variety of disparate terms makes consistent analysis of learning activities difficult.

The eDAT, as described in the following, utilises the two commonly used terms (‘interaction’ and ‘feedback’) that are associated with higher retention in ODL. The consistent use of these terms, as suggested by the following analysis, could enable a more accurate and effective way for tutors and learning designers to describe learning activities. When learning activities can be accurately described, they can be quantified and used with learning analytics to provide evidence for effective learning designs that increase retention (Bakharia et al. 2016).
Methods
The literature discussed suggests that retention is increased when ODL includes interaction and feedback activities. However, these terms may not be used by tutors in the same way. These terms were tested using content analysis methodology to identify the extent to which tutors were using them consistently.

Content analysis
Content analysis is a method of quantifying text to enable statistical analysis of the text by a process of ‘coding’ or categorising. It is a ‘research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use’ (Krippendorff 2013, p. 24). It has been used in a variety of educational settings, for example to analyse the impact of tutors’ roles in online discussions (Dubuclet, Lou, and MacGregor 2015). To carry out a valid and reliable content analysis for this study, the following steps were taken:

1. specifying the units of analysis
2. identifying learning activity vocabulary to test
3. recruiting raters
4. calculating inter-rater reliability (IRR)

(adapted from Neuendorf 2002, p. 50)

Specifying the units of analysis
For this study the specific learning activities or task descriptions written by tutors and presented in the VLE for students were analysed. A convenience sample of four distance learning modules from one Higher Education (HE) institution were chosen to represent a variety of courses. They were varied and from different subject areas (law, politics, games and sport), aimed at different levels (undergraduate and postgraduate) and included a total of 215 learning activities of different types and lengths.

Identification of units of analysis is critical but also challenging (Gorsky and Blau 2009). If the unit of analysis is too general, it may be easy to categorise but hard to analyse; if too small, it may be difficult to categorise reliably. For this study the units of analysis were prepared by splitting activities into multiple parts based on the learning activity ‘verbs’. For example, a typical student activity was as follows:

1. Read xx, answer the following [structured] question and then post your response to the forum.

This was divided into the following for analysis:

(i) read xx,
(ii) answer the following [structured] question and then
(iii) post your response to the forum.

Some courses included ‘optional activities’, for example, extended reading or open forums. These were also included as units of analysis because the impact of voluntary participation may be significant (So 2009).
Identifying learning activity terminology to test

Based on the literature mentioned, analysis was conducted on the learning activity terms ‘interaction’, ‘feedback’ and ‘other’. Activity types and examples were provided to assist the rater when categorising each activity, as in Table 2.

Table 2. Terminology tested.

<table>
<thead>
<tr>
<th>Activity terminology</th>
<th>Activity type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interaction</strong> with ...</td>
<td>A. the tutor</td>
<td>online webinar/lecture, 1–1 tutorial, coaching session, email, phone</td>
</tr>
<tr>
<td></td>
<td>B. other students</td>
<td>forum discussion (may include tutor), group work, peer assessment, adding comments to peer wikis or blogs</td>
</tr>
<tr>
<td></td>
<td>C. (interactive) content</td>
<td>computer simulation, multimedia interactions (excludes interaction with text or video)</td>
</tr>
<tr>
<td><strong>Feedback</strong> from ...</td>
<td>1. the tutor</td>
<td>formative or summative feedback or grades</td>
</tr>
<tr>
<td></td>
<td>2. peers</td>
<td>structured peer-assessment exercise, grading activity</td>
</tr>
<tr>
<td></td>
<td>3. self-feedback</td>
<td>using model answers, self-reflection, trial and error exercises</td>
</tr>
<tr>
<td></td>
<td>4. computer (automatic)</td>
<td>from computer simulation, computer-marked test</td>
</tr>
<tr>
<td><strong>Other</strong> activities</td>
<td></td>
<td>reading or watching, research, creating</td>
</tr>
</tbody>
</table>

Recruiting raters

In many studies, only two raters are used when a larger number would produce greater validity. Independent raters may be unbiased, but in many studies raters are either researchers or the researchers’ assistants (e.g. Rienties and Toetenel 2016). Raters require familiarity with the language and context for analysis but not overfamiliarity with specialised vocabulary, which may reduce the universality of their analysis (Krippendorff 2013). Here, all four raters were academic colleagues, familiar with educational terminology and who completed the content analysis task independently following training.

Calculating inter-rater reliability

When raters all agree, this increases confidence that the analysis is consistent and objective and that other raters would be likely to obtain the same result. However, even high reliability scores do not guarantee validity. For example, raters may all display the same prejudice or use the same concepts as others in a specialised community. High reliability may also indicate a loss of validity; for example, the categories may be oversimplified or superficial (Krippendorff 2013). In addition, high agreement between raters may simply mean that a particular item is missing from the content being analysed or that there is a high degree of similarity between the items being rated.
Inter-rater reliability is often measured using Cohen’s kappa, but this has been criticised as it encourages the use of just two raters when more raters would provide more robust findings (Krippendorff 2013). Krippendorff’s alpha ($\alpha$) is a more effective measure of IRR as it can be applied to any number of observers, any number of categories, any metric or level of measurement, as well as to incomplete data and large and small sample sizes (Krippendorff 2011), and has been used in this study to calculate IRR.

There is no statistical rationale presented in the literature for acceptable levels of IRR. Krippendorff (2004) suggests that where the analysis is critical, a level of $\alpha \geq 0.800$ should be considered necessary, and in situations where conclusions may be more tentative an IRR of $\alpha \geq 0.667$ may be acceptable.

All 215 learning activities from four courses were categorised by four raters. Each course was rated independently, and each activity was categorised as ‘interaction’ and/or ‘feedback’ or ‘other’.

Results

The raters’ overall categorisations of ‘interaction’ or ‘feedback’ for each activity were compared, and IRR was calculated with Krippendorff’s alpha. There was some disagreement among raters and although the ‘interaction’ category had an acceptable level of agreement, the ‘feedback’ categorisations were near to but did not reach an acceptable level of IRR as in Table 3.

Table 3. Inter-rater reliability results.

<table>
<thead>
<tr>
<th></th>
<th>‘Interaction’</th>
<th>‘Feedback’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-rater reliability: 4 raters and 215 activities</td>
<td>0.815</td>
<td>0.612</td>
</tr>
</tbody>
</table>

Discussion

The IRR figures show the difficulties in categorising learning activities even when using the commonly used terms ‘interaction’ and ‘feedback’.

In total, of the 308 possible discussion-type categorisations, 285 were categorised as peer interaction and 197 as peer feedback. A significant issue was the way discussion forum activities were written; for example, discussion-type activities included five different terms: ‘discuss’, ‘post’, ‘comment’, ‘post & comment’ and ‘post & discuss’. Raters categorised both ‘discuss’ and ‘post’ activities as including feedback when this was not indicated in the task. Sixteen ‘discussions’ were rated as peer feedback. In addition, discussion activities were sometimes categorised as ‘other’, perhaps because raters thought that posting on a forum did not comprise interaction. Within this variety of categorisations there was also noted a lack of consistency within raters. The highest level of agreement was for activities that specified both ‘post/comment’ and ‘post/discuss’, suggesting a greater clarity in the task.

There were noticeable differences between raters when categorising feedback activities. For example, one rater categorised the activity ‘Students access Blackboard
for topic lecture notes, videos, etc. Try to apply these techniques to your own work’ as feedback when no other rater had categorised it as such. Another rater categorised the activity ‘Please post … on the discussion board’ as feedback 22 times when the other raters did not. Assessment activities were not consistently categorised as feedback, presumably because this was not specified in the activity.

Some learning activities that were inconsistently categorised did not conform to good practice recommendations for interaction activities (e.g. Akin and Neal 2007; Salmon 2004) or recommendations for feedback (e.g. Nicol and Macfarlane-Dick 2006). However, a good practice example – ‘Students post questions/comments in bulletin board for peer and tutor discussion’ – was categorised the same way by all raters.

The selection of courses for this study included a variety of subject disciplines, and the raters were from different disciplines. This may have impacted on the ways the learning activities were written and also on the individual ways that the raters interpreted both the learning activity and the terms in the eDAT when completing the content analysis task. Further research in this area is needed.

Conclusion

Feedback on the effectiveness of learning designs is needed to improve ODL, but this is difficult to obtain without a consistent way to describe learning activities. Two types of activities are highlighted in the literature as having the potential to improve retention and quality of online learning: interaction with tutor and peers and feedback on learning. However, despite these terms being commonly used, they were difficult to apply consistently to the learning activities in this study. The eDAT utilises this terminology to help improve categorisation and quantification.

The difficulties in using common terms to categorise learning activities was surprising. The IRR for interaction was acceptable, but the IRR for feedback did not reach an acceptable level, suggesting that this is a complex term, difficult to use consistently. These terms, as used by the online course designers and by the raters, have different implicit meanings and reflect different teaching perspectives (Trigwell, Prosser, and Ginns 2005). However, the example given of an activity categorised consistently suggests that increased clarity about opportunities for interaction and feedback in a task will improve consistent use of these terms.

The eDAT has been developed to attempt to address these issues. It builds on the other Learning Design representation tools mentioned but focusses on two key online learning activities that are associated with higher retention. The eDAT enables tutors and designers to carry out the analysis described, that is, to categorise their learning activities using the terms ‘interaction’ and ‘feedback’ and to quantify them. Interaction activities can be categorised with some confidence, but feedback activities may be less easy to identify and require review and editing for clarity. Further analysis of the effectiveness of the tool is being conducted and will be reported separately.

Using the eDAT to categorise learning activities helps to provides quantitative data about the learning design. It also highlights to tutors the need to specify clearly to students when and how they will be interacting with others and when they can expect to receive feedback on each of their activities, thus potentially improving the learning design.
Appendix 1: The E-Design Assessment Tool

The E-Design Assessment Tool (Walmsley 2017) employs the tested terminology in both a Word template and Excel for use by tutors and designers, together with examples and a guide to quantifying learning activities. A sample follows, and both are freely available for download from the eDAT site: http://blogs.staffs.ac.uk/bestpracticemodels/edat/.

E-Design Assessment Tool

Tutor instructions: Add your activities below and indicate where you have specifically included interaction and/or feedback activities. Calculate the % of each activity type to help you reflect on your learning design. Use retention and attainment rates to evaluate the quality of the learning design.

<table>
<thead>
<tr>
<th>No</th>
<th>Specific learning activities/tasks (you may need to split activities that include separate parts)</th>
<th>Interaction with…</th>
<th>Feedback from…</th>
<th>Other content or activities ✔</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A. tutor</td>
<td>1. tutor</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B. peers</td>
<td>2. peers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C. (interactive) content</td>
<td>3. self</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. computer (automatic)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Activity text here …</td>
<td>[add interaction type here if present in activity]</td>
<td>[add feedback type here if present in activity]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Insert additional rows as required]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total activities: __ % with interaction % with feedback % other</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


