Exploring design features to enhance computer-based assessment: Learners’ views on using a confidence-indicator tool and computer-based feedback

Ingrid Nix and Ali Wyllie

Ingrid Nix and Ali Wyllie are Lecturers in Learning and Teaching Technologies in the Faculty of Health and Social Care at The Open University, and Teaching Fellows in the COLMSCT CETL, researching e-learning design improvements and learner experiences of computer-marked assessment. Address for correspondence: Ms. Ingrid Nix or Ali Wyllie, Horlock 222, The Open University, Walton Hall, Milton Keynes, MK7 6AA, England. Tel: 01908 654124; Email: i.nix@open.ac.uk or a.j.wyllie@open.ac.uk

Abstract

Many institutions encourage formative computer-based assessment (CBA), yet competing priorities mean that learners are necessarily selective about what they engage in. So how can we motivate them to engage? Can we facilitate learners to take more control of shaping their learning experience? To explore this, the Learning with Interactive Assessment (LINA) project, a small-scale study, trialed a selection of online, interactive question features designed to increase motivation, particularly of work-based learners, and to enhance learners’ ability to self-regulate their own learning experience. We present initial findings on learner perceptions of: (1) a confidence indicator tool—for learners to indicate their confidence that their answer will be correct before answering a question; (2) a learning log—for learners to insert reflections or view system-recorded information about their learning pathway; and (3) question feedback—for displaying a range of author comments. Themes emerged relating to motivational, behavioural and cognitive factors, including risk taking and skills in self-assessment. A serendipitous finding highlights learners’ frequently not reading onscreen feedback and presents perceptions on when feedback becomes most useful. Motivators and barriers to engagement and self-regulation are discussed in relation to principles of best feedback practice.

Introduction

Considerable progress has been made in understanding what makes for effective feedback in written assignments. Nicol and Macfarlane-Dick (2006) identified seven principles of good feedback practice that support learner self-regulation. Gibbs and Simpson (2004) and Hounsell (2007), among others, have shown that students are strategic...
and selective about their use of formative feedback. But while opportunities have been created for more frequent and effective feedback, what motivates learners’ engagement in formative assessment?

The affordances of online assessment make it possible to offer students’ choice on entering the assessment environment, placing them in a role where they can configure and control their own mode of learning and resulting learning experience.

In the Faculty of Health and Social Care, we have used CBA for social work students to develop European Computer Driving Licence equivalent skills, and for nurses to review course concepts. However, we wanted to encourage greater engagement, as our students typically lack confidence in using computers.

The resulting Learning with Interactive Assessment (LINA) project aimed to give learners increased control of their learning experience.

Research questions
What features would enhance learners’ self-regulation and motivation to engage?

Would a confidence indicator tool and a learning log promote a more reflective approach to learning?

Figure 1: LINA confidence-indicator tool and animation open

© 2009 The Authors. British Journal of Educational Technology © 2009 Becta.
LINA overview
LINA uses an existing assessment system (OpenMark), introducing new and enhanced features. After viewing a help sheet introducing LINA features, students choose a topic: creating a table or searching for information. Each topic contains a sequence of 10 questions, six formative (L1–L6) and four summative (T1–T4), linked by a narrative-scenario relevant to their work practice.

Students select formative questions from the practice sequence (Figure 1), progressing towards readiness to do the test questions. Media resources contain task-related information. At any point, students can open the learning log (LL) to review their actions and score or enter reflections. After seeing the answer options and before submitting their first attempt, the student indicates their confidence level using the confidence-indicator tool, whose setting affects their score. Questions allow three attempts. Each failed attempt gives increasingly more feedback. The student can skip or revisit any practice question but can submit test questions once only.

Methods
Research instruments
We gathered and triangulated data using:

1. A website introducing LINA—two topics of 10 questions each.
2. A feedback questionnaire containing statements with a 5-point Likert scale, and open-ended text input questions.
3. Video of three participants testing LINA in a ‘Userlab’ facility, and using a Think-Aloud protocol to talk through their actions (produced video transcripts).
4. A computer-generated learning log of participants’ actions and pathways, plus additional text reflections entered by participants.
5. Notes taken by authors during videoing.
6. Audio interviews following Userlab testing (produced audio transcripts).

Sample
Twelve social work students from two cohorts of a second-level course volunteered to trial LINA online over 4 weeks in 2007 and 2008. All had previous experience of CBA using OpenMark.

The three Userlab participants were a social work student, referred to as ‘S’, and two nursing staff-members, ‘C’ with previous CBA experience, and ‘D’, studying an Open University course.

Approach
During Userlab recordings, the authors noted additional factors, such as eye direction and body language. We compiled questions based on our observations, put to each
participant in follow-up interviews. Emerging themes from the three audio interviews were triangulated against data from the feedback questionnaires of the 12 students, the video transcripts, systems data and reflections captured in the learning logs. Gender-related analysis was not considered due to the sample’s limited size.

**Results from the study**

*Findings: confidence-indicator tool*

Building on the work of Davies (2005) and Gardner-Medwin (2006), a confidence-indicator tool was included to encourage learner reflection. Participants first engage with the question and select their answer before indicating their certainty level that their chosen answer is correct. The confidence-indicator tool offers three options: high/medium/low (Figure 1).

Learner choice affects the score for each item (Table 1). Over-confidence combined with an incorrect answer results in negative marks to discourage guessing and encourage reflection. Learners draw on their own evidence-base as justification to improve their reflective practice, a key requirement in social work and nursing professions.

<table>
<thead>
<tr>
<th>Confidence level setting</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attempt 1 correct</td>
</tr>
<tr>
<td>Low</td>
<td>2</td>
</tr>
<tr>
<td>Medium</td>
<td>3</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
</tr>
<tr>
<td>None selected</td>
<td>-2</td>
</tr>
</tbody>
</table>

**Discussion**

Gibbs and Simpson (2004) have argued that assessment works best to support learning when a series of conditions are met. In our findings did we see evidence of these conditions and the feedback principles of Nicol and Macfarlane-Dick (2006).

The quantitative analysis of respondents’ views of the confidence-indicator tool revealed that 83% felt it might improve their reflective practice, and 92% thought that the marks were fair. Students stated that they valued the score as well as the feedback, but from the qualitative data, a more complex story emerges. Only 33% valued the score as a reflection of achievement, the others perceived the value of the confidence-indicator tool to be a prompt to reflection and justification of their choice: ‘... it makes you think about what you’ve just said ... It makes you evaluate ... how well you think you’ve done’ (D-78).
Is the score important? For some, it is a reflection of achievement, but for others, the score value itself is not as important as the encouragement it engenders (Principle 5). D suggested more meaning would result if the scores were: ‘registered against something as a means of evaluating how you’d done overall’. S suggested personalisation of the score to improve motivation: ‘if I saw little gold stars or something ... I think that would motivate me’ (S-188). C felt the score encouraged more engagement even when he already felt confident in the skill: ‘it just introduces a bit of an edge to it in a gambling sort of sense’ (C-98).

An important but unexpected factor that emerged was that not all learners were good at self-assessment, misjudging their abilities in relation to questions and thereby undermining their scores by selecting lower confidence levels than was reflected in their results. Table 2 shows learners’ confidence level for each question and at which attempt they got the answer correct (or not).

In Topic 1, C’s confidence rating (High) was well judged in only 75% of questions, and he admitted ‘gambling’ by choosing high confidence to maximise scores. In Topic 2, he showed high confidence in 90% of questions, well judged in 100%, having lowered his confidence to Medium in Question 4, which was subsequently incorrect. Therefore, when an incorrect answer occurred, it was always when a High rating had been set, suggesting a surface approach to learning.

In contrast, D often selected medium or low confidence, suggesting she would need several attempts to get the question correct, but in the majority of cases answering correctly on the first attempt. Since her confidence-rating was well judged only 50% of the time in Topic 1 and 70% in Topic 2, her inability to assess her knowledge, or lack of confidence in that knowledge, lowered her score.

Although stating she was not fully aware of negative marking during Topic 1, D continued to misjudge her ability in Topic 2, but was less concerned about the result stating: ‘It wasn’t so much about the score, it was about working out where you’ve got it wrong’ (D-64). For her, self-assessment encouraged reflection and deeper learning, fulfilling Nicol’s second principle (facilitating self-assessment), in which ‘by having to rate their confidence students are forced to reflect on the soundness of their answer and assess their own reasoning’ (Nicol, 2007: 58).

Confident participant C confirms: ‘... when you’re highly confident and then you don’t get it right, that inevitably makes you stop and think’ (C-104). However, not everyone agrees on the value of self-assessment. Participant S felt: ‘it’s not for me, it’s for you, you’re teaching me, it’s for you to say how good I am at this’ (S-145).
Findings: learning log

A condition for formative feedback is that students should be able to monitor their work as they proceed through the assessment task in order to improve (Sadler, 1998: 77). Whereas OpenMark currently does not provide systems data (actions and pathways) to

<table>
<thead>
<tr>
<th>Question</th>
<th>Practice 1</th>
<th>Practice 2</th>
<th>Practice 3</th>
<th>Practice 4</th>
<th>Practice 5</th>
<th>Practice 6</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
<td>Confidence: High</td>
</tr>
<tr>
<td></td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
<td>Correct at attempt: 1st</td>
</tr>
<tr>
<td>Practice 1</td>
<td>Topic 1</td>
<td>Topic 2</td>
<td>Topic 1</td>
<td>Topic 2</td>
<td>Topic 1</td>
<td>Topic 2</td>
<td>Topic 1</td>
<td>Topic 2</td>
<td>Topic 1</td>
<td>Topic 2</td>
</tr>
<tr>
<td>Practice 2</td>
<td>No attempt</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Practice 3</td>
<td>High</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Practice 4</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Practice 5</td>
<td>No attempt</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Practice 6</td>
<td>High</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Test 1</td>
<td>High</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>No attempt</td>
<td>No attempt</td>
</tr>
<tr>
<td>Test 2</td>
<td>High</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>No attempt</td>
<td>No attempt</td>
</tr>
<tr>
<td>Test 3</td>
<td>High</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>No attempt</td>
<td>No attempt</td>
</tr>
<tr>
<td>Test 4</td>
<td>High</td>
<td>High</td>
<td>1st</td>
<td>1st</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>No attempt</td>
<td>No attempt</td>
</tr>
</tbody>
</table>

Summary of confidence

<table>
<thead>
<tr>
<th>Summary of confidence</th>
<th>Participant C</th>
<th>Participant D</th>
<th>Participant S</th>
</tr>
</thead>
<tbody>
<tr>
<td>High 100%</td>
<td>High 90%</td>
<td>High 70%</td>
<td>High 90%</td>
</tr>
<tr>
<td>High 100%</td>
<td>High 90%</td>
<td>High 70%</td>
<td>High 90%</td>
</tr>
<tr>
<td>High 100%</td>
<td>High 90%</td>
<td>High 70%</td>
<td>High 90%</td>
</tr>
<tr>
<td>High 100%</td>
<td>High 90%</td>
<td>High 70%</td>
<td>High 90%</td>
</tr>
<tr>
<td>Summary of attempts</td>
<td>1st 75%</td>
<td>1st 90%</td>
<td>1st 90%</td>
</tr>
<tr>
<td>Correctly judged</td>
<td>75%</td>
<td>100%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Note on data quotations: eg, S-5, D-22, C-1 (the letter refers to a participant, the number refers to the transcript section number).

Topic 1–creating a table; Topic 2–searching for information.

© 2009 The Authors. British Journal of Educational Technology © 2009 Becta.
students, our approach was to make this available during the question-answering process. In addition, we enabled students to input their reflections (Figure 2) at any point.

Questionnaire results revealed that 66% of respondents found the log useful. But how was it used? Testers were automatically prompted for reflections before leaving each question. They input comments covering: level of question difficulty, lack of clarity in question wording, comments on errors, action points, reasons for choices and feelings.

However, reviewing their log was entirely voluntary. There were indications that familiarisation was required to adjust to this proactive role. Having started by using it retrospectively at the end of Topic 1, D switched to using it concurrently, throughout Topic 2, noting the flexibility to ‘pop across, in and out as I wanted to’ [D-5] and confirming she might use both approaches in future. S commented that the newer the learning, the more she would enter into the log, suggesting she recognised the potential value of self-generated feedback.

So how was this feedback used? Although the log records the score, the testers did not pay close visual attention to this information. The log was perceived less as a motivational tool than a way to retrace actions, monitor progress, and support pragmatic

![Figure 2: Learning log example](image-url)
follow-up. S was more interested in refreshing her memory, reviewing her reflections: ‘I
noted “score” but I was reading the bits [about the misspelt word] of the learning log so that’s
what I was concentrating on’ (S-7).

Interestingly, two of the three participants commented that they would be more likely to
use the log after an incorrect answer, suggesting that systems data and personal feed-
back becomes most useful for Principle 6: ‘closing the gap’. S and D adopted a similar
strategy. D commented: ‘... if you’ve got something wrong it makes you ... address it and look
at it again, and that’s how you learn’(D-2).

So while the log supports self-regulation, how did it engage learners? D appeared to find
it reassuring to know it was available: ‘I think if you know you’ve got that facility to
monitor as you’re going along ... it’s useful’ (D-8). This suggests that having information at
one’s fingertips fostered feelings of being in control. S gave further ways she might use
the log: to review information outside the assessment environment and to query dis-
puted answers with the course team. Finally, she would like the system to prompt her,
on her next visit, with previous reflections on her incorrect answers.

Discussion
These findings demonstrate that participants valued the combination of computer-
generated system information that recorded their past decisions, along with their
self-generated content of typed reflections. Having this easily retrievable facilitated
self-monitoring and self-regulation across a number of criteria, using different types of
information covered within the log. Thus, like the confidence-indicator tool, it addresses
Nicol and Macfarlane-Dick’s (2006) Principle 2, supporting self-assessment and reflec-
tion. Testers appear to find the log most useful when they identify they need support,
notably after an incorrect answer. They may then decode their mistake so as to improve
their performance (Principle 6: ‘closing the gap’). Knowing they have access to a log
builds confidence, aware that they can refer to it when necessary.

However, testers found it more challenging to type their own comments into the log. Our experience showed that they needed prompting with examples, such as reflections
on errors, or reminders of feed-forward comments. Less experienced students may not
have the self-awareness to be able to provide reflections that are worth revisiting. This
feature puts students in control of potentially providing high-quality feedback for them-
selves. Such a level of control and responsibility may suit more confident and experi-
enced learners who can then use it to self-assess and self-correct. This ‘delivers high
quality information to students about their learning’ (Principle 3), but untypically for
CBA, responsibility for the quality of feedback lies in the hands of the learner rather
than the teacher.

Findings: question feedback
Whereas the learning log provides system and self-generated ‘internal’ feedback, ques-
tion feedback provides ‘external’ pre-authored feedback, specifically to provide oppor-
tunities to close the gap (Principle 6). LINA questions allow three attempts with
increasingly supportive feedback after each attempt (Figure 3). The final feedback may include: a notification of correct/incorrect answer, why it was incorrect, what the correct answer was, an explanation, additional tips and a reference to relevant course materials. These cover recognised stages of feedback (Gibbs & Simpson, 2004; Miller, 2008).

So why did testers frequently not refer to this feedback? C reported that he did not use feedback when he was clear what the mistake was, closing the gap himself. Furthermore, even when unclear, he stated: ‘I thought ‘I’m wrong, what did I do wrong’, and then I go back myself and try and analyse it myself to find out what went wrong’ (C-3). Describing himself as a confident learner, C had what he called a five-step strategy for closing the gap, involving rechecking the question resources for the answer.

When C saw an answer was correct, he reported that feedback had no value. C appears achievement oriented: ‘Whatever the situation, if the end result is that I’m going to get a score, and that’s going to make a difference to, for example, my course result, that’s going to be a very important driver for me’ (C-52). He tends towards risk taking to get the maximum score. As a result of an achievement focus and a confident strategy, for C, feedback has little perceived value.
Interestingly, ‘Towards the end there I did read some feedback because I recognised that I’d been paying no attention to that up to that point’ (C-1). This conscious effort to interrupt the pattern of his responses and appraise the assumptions on which it was based suggests a self-reflective learner, capable of critiquing his self-regulatory practices.

By contrast, D did not start with the same confidence in her management of the environment. As with the learning log, she developed her use over time. Likewise, S developed distinctions in her use. Both their approaches were influenced by being learning oriented, and describing themselves as ‘methodical’ and ‘apprehensive’ (S3-55), and ‘systematic’ and tending ‘not to cut corners’ (D-24), they adopted strategies to suit these learning approaches.

Their attitude towards feedback depended on their perceived readiness to take on board new or additional information. S indicated she would be receptive during formative new learning, seeing feedback as playing an instructional role and providing different perspectives and transferability (Lizzio & Wilson, 2008): ‘when I got that question wrong I could then read and think that’s why, that was where you were coming from, and it was different to where I was coming from’ (S-2). At such time, they also valued tips ‘setting out other options’ (D-5) to extend their knowledge. However, in order to facilitate such selective use, S suggested such tips be made distinguishable from general feedback, eg, in colour, reflecting her tendency to: ‘...highlight all the [tips] so that they stand out to me’ (S-8).

S and D were least receptive to feedback when the answer was correct, including knowing their score. D stated: ‘I would look on it as a learning thing, the fact that I’d actually worked through and got the question right at the end’ (D-3). S would avoid further feedback when less confident, especially in summative assessment, as she found it confusing to read additional methods of doing the same task: ‘...as long as the way I’ve done it gets a correct answer, that’s all I’m interested in’ (S-4).

Discussion
Our findings reveal that at the time of assessing their need for feedback, participants who are learning oriented (personal satisfaction driven) and confident are more ready to read and act on feedback—the feedback is valued because it can either help close a gap or extend learning. They appear to relate to the feedback as multi-dimensional (Miller, 2008) due to the different components within it. Familiarity with the environment is required to encourage use.

However, if someone is achievement oriented (score-driven) and confident, they may prefer personal strategies and ignore feedback—the feedback is not valued because it is not needed to help close the gap, and additional learning is not sought.

Concerning the range of feedback provided, fuller feedback following a correct answer may not be appropriate, since it was often ignored. Miller (2008) has noted the need for further studies of students’ perceptions of what constitutes useful feedback in CBA. As
she noted, students may decline to access feedback and use strategies and features within the assessment environment so that they ‘circumvent learning’. When their motivation is driven by score rather than understanding of a concept, she suggests that students may be focusing on the summative rather than formative purpose of CBA. Our findings suggest that students would value more personalised control over feedback to help make selective use of it while judging its timing, quality and relevance.

**Conclusion**

The findings suggest that learners respond variously to reflection tools linked to their learning approaches. The confidence-indicator tool and learning log were motivators for engagement in learning and assessment in different ways to different learners. Learners welcomed the opportunity to configure and control their own mode of learning, though they needed practice in the skills of self-regulation.

Pausing to decide confidence and justify decisions encourages self-assessment and reflection and may encourage higher-order cognition, as identified in Nicol and Macfarlane’s model of self-regulated learning (Nicol & Macfarlane-Dick, 2006). While not all students are good at self-assessment, they may still benefit from the prompted reflection.

Our findings suggest that confidence-based self-assessment may be more relevant during formative than summative learning. Although participants did not pay close visual attention to scores, the confidence-indicator tool may have induced achievement-oriented strategies (Lizzio & Wilson, 2008), which may drive learners away from using formative feedback (Miller, 2008).

The confidence indicator motivated learners to engage even when confident of their knowledge base through increasing the element of score-related ‘risk’. Unconfident learners might be less averse to risk if the score could be personalised, eg, as applause, and thereby appear to reward rather than penalise.

The learning log proved valuable in supporting self-regulation by enabling flexible access to varied feedback, controlled by the learner. The nature of the comments typed into the log indicated different perceived audiences and reach. This compares favourably with the limited control of question-based feedback. Whereas the log enabled self-generated input, question feedback contained feedback categories determined by the author, not learner. Therefore, to best support learners, author-designers should fully exploit all those capabilities within feedback features, which enable learners to make optimum choices.

**Implications for practice**

Institutions are developing virtual learning environments to provide a range of integrated features including, eAssessment. These tools need to be fit-for-purpose and to accommodate different learning needs, and learners need to develop expertise in using
the tools effectively. The more we can prepare and develop learner skills in managing their own learning environment, the more we will be equipping them to be effective learners.

By allowing choice within the assessment environment, we liberate learners to do things their way, reflecting their characteristics, and harnessing their motivation to achieve the best result.

Limitations and future research
Although the participant sample was small and not fully representative of the wider student population, this study has enabled some interesting findings to emerge that call for further investigation. Future research is needed to explore the relationship of learner personality characteristics to their likely use of feedback (Lizzio & Wilson, 2008), in particular how perceived control over learning or achievement-oriented strategies might influence use of feedback.

Acknowledgement
The LINA project is funded by the Centre for Open Learning of Mathematics, Science, Computing and Technology Centre for Excellence in Teaching and Learning at The Open University.

References