

Online MCQ - formative or summative assessment? Experiences from a higher educational environment.

Monica Johannesen

Faculty of Education and International Studies
Oslo University College, Norway
monica.johannesen@lui.hio.no

Laurence Habib

Centre for Educational Research and Development
Oslo University College, Norway
laurence.habib@hio.no

Abstract: Within the field of Higher Education there has been a growing interest in formative assessment. A range of online tools for multiple choice questions (MCQ) are often made available to academic staff, but the extent and level of use of such tools vary greatly. This paper is based on an empirical investigation of actual use and attitudes of lecturers towards a particular MCQ tool both for formative and for summative assessment. We use Draper's categorization of MCQ use as a conceptual lens for the study. The research methodology is interpretive, with interviews and diaries forming the main source of empirical data. The findings from the data set suggest that many lecturers are openly resisting the tool, while those who do use the tool do so predominantly to support formative assessment and thereby encourage their students to develop higher-order learning skills.

Introduction

According to Black (Black, et al., 2003) the purpose of assessment is threefold: making schools accountable, providing students with certificates and structuring their learning. During the last few years, the theoretical field of formative assessment has drawn considerable attention (Black, et al., 2002, 2003; Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006; Sadler, 1998) and assessment is now widely recognized as a powerful factor in promoting learning.

Formative assessment has a strong standing in Norwegian education, although assessment is also widely used summatively. Virtual learning environments (VLEs) have been introduced at all levels of the Norwegian educational system, providing a range of tools to support both formative and summative assessment. Among them figure a range of tools for creating multiple choice questions (MCQ), thereby facilitating the task of creating tests for students. MCQ tests have been described as encouraging the reproduction of factual knowledge and promoting rote learning and superficial learning (Gibbs, 2006). For that reason, multiple choice assessment (MCA) tools tend to be regarded as incompatible with formative assessment (Draper, 2009a).

An informal survey of colleagues from several Norwegian institutions of higher education suggested that VLE-based MCA tools were used relatively little. This prompted us to investigate whether this was due to a negative attitude towards such tools and the pedagogy they represent, to the teachers' lack of competence in using the technology or to other factors. In this paper, we aim to uncover how academic staff members relate to MCQs as formative and summative assessment tools.

E-assessment

A recurring theme in the literature on computer-based assessment systems (e-assessment) is the risk that such systems will encourage the test designers to favor things that are easy to measure rather than things that are harder to assess (Ridgway, et al., 2004). This suggests that the use of e-assessment requires a high awareness of what the subject of the assessment actually is.

The relationship between information and communication technologies (ICT) and assessment is multifaceted. At least three different perspectives can be identified on the topic (Erstad, 2009; McFarlane, 2003; Ridgway, et al., 2004). Computer-based assessment tools can be used to assess traditional skills and can build upon prior knowledge in assessment. Although such testing can be based on existing tests, new challenges emerge. There is evidence of under-performance among students that have no or little computer proficiency when answering a computer-based test as well as under-performance among computer literate students when using paper-based tests. (Russell, et al., 2003). It is interesting to note that even "complex" knowledge can be the object of computer-

supported assessment, which suggests that there might be a high degree of correlation between human and computer-based marking of essays. Studies conducted by, e.g., Jordan & Mitchell (2009) and Landauer, et al. (2003) provide evidence that computer marking is often close to human marking. This indicates that computer-based assessment may be valuable for more than surface learning (Gibbs, 2006)

The use of ICT in assessing new educational goals like metacognition, creativity, group projects and communication skills is also of interest to the field. Such higher-order skills are difficult or even impossible to assess using traditional methods. Computer-based systems are essential in providing realistic problem solving and simulation of real-life situations. However, they raise new problems in terms of assessment, such as how to assess the dynamic aspects of problem solving (Wirth & Klieme, 2003) and how to separate problem solving from related knowledge (Bennett, et al., 2003). According to Ridgeway & McCusker (2003) and Shepard (2009) there is a need to clarify the role of ICT in assessment. Ridgeway & McCusker (2003) suggest that an assessment framework needs to be developed and aligned with earlier assessment systems so as to function as a key change agent.

Finally there is a need to take seriously the issue of digital literacy among the students themselves. ICT proficiency is essential for much modern living, and, in that respect, there is a need to develop assessment forms that will encourage students to increase their level of digital literacy (Sieber, 2009). When attempting to solve complex problems, students will have to rely on a combination of generic computer skills, networking strategies and reasoning. This combination of skills and knowledge is crucial for modern living and therefore needs to be reflected onto the curriculum (Quellmalz & Kozma, 2003).

E-assessment for learning with MCA tools

Formative assessment is defined as "a process, one in which information about learning is evoked and then used to modify the teaching and learning activities in which teachers and students are engaged" (Loddington, et al., 2009: 122). It is therefore crucial that technology-enhanced assessment is designed and used for purposes that support such iterative feedback processes.

It has been suggested that MCQs may support deep learning and higher-order thinking if used in certain ways (Draper, 2009a). Techniques such as using assertion-reason questions, letting the learner generate reasons for and against each response option, and allowing confidence-based marking have been tried with good results in terms of increased incidence of deep learning. Brain-teasers aimed at prompting peer discussion and letting students creating their own MCQs have also been successfully used to support deep learning.

There is a growing interest in using technology to support formative assessment that is meant to foster higher order learning (Draper, 2009a, 2009b; Heinze & Heinze, 2009; Shephard, 2009). In particular, Draper (2009a) suggests that the use of MCQ, although it is often seen as promoting surface learning, may also be used to advance deep learning (2009a:290). Based on the work on peer instruction and peer interaction from Mazur (1997) and Howe et al. (2005), he proposes a categorization of possible learning designs using MCQs to achieve deep learning. He identifies six learning design categories:

1. **Assertion–reason questions**, that ask directly about which theories or reasons can be linked to particular facts.
2. **Reason generation**, i.e. having the learner take an MCQ and generate reasons for each alternative answer, i.e. an explanation of why each option is a suitable answer to the question or not.
3. **Confidence-based marking**, i.e. having learners both choose an answer to each question and state their level of confidence for that answer.
4. **Brain-teasers to prompt peer discussion**, i.e. MCQs that are designed to create uncertainty both at the individual level and at the group level, so as to foster discussion.
5. **Having students create MCQs and present them** to the rest of the class, thereby bringing about discussions about the process of constructing the questions.
6. **Having students create MCQs that will be used as formal assessment forms** for the rest of the class. This includes presenting justifications for each response option.

In our case study, we have attempted to investigate the variety in attitudes towards the possible role of MCQs in the design of learning. The data gathered bears much relevance to many of the categories listed above, and we therefore propose to use Draper's categorization to structure our presentation of the data.

Methodology

This paper is based on a study of the use of VLE-based multiple-choice assessment tools in a Norwegian institution of Higher Education. In this study, we focused on gathering information on patterns of use and attitudes toward a particular VLE, which was a commercial software acquired by the central management of the institution and made available to both staff members and students. The investigation was carried out as an explorative case study, whereby VLE usage was explored within the broader context of teaching practice. Interviewees were selected using purposive sampling (Miles & Huberman, 1994; Patton, 1990), using both online ads and recommendations. All of them were academics involved in core teaching activities and users of the institution-wide VLE.

The empirical data base consisted of a combination of face-to-face interviews (each lasting about one hour) and personal diaries kept by the participants over a period of one week, providing a detailed overview of their VLE-based activities, their aims and their place in the teaching practice of the diary keeper. The study was conducted longitudinally so as to uncover changes that may have occurred over time in terms of actual use and attitudes. All the collected diaries and the transcribed interviews were loaded into a computer-assisted qualitative data analysis software, and coded according to a list of keywords developed by the authors. The extracts presented below were located via searches using keywords that were deemed relevant to the topic of this paper.

Findings

The first two categories presented in Draper's article (*assertion-reason questions* and *reason generation*) focus specifically on avoiding surface learning. However, our data suggests that such categories do not have direct relevance to the daily practice of lecturers. On the contrary, the lecturers express concerns that MCQs would promote surface learning. One concern is that multiple-choice assessment is considered more appropriate for subject areas where there is a common agreement as to the existence of "exact knowledge".

Lecturer: But I think that, in a way, I think that multiple choice assignments have a function. Or, that is, that they can be just as nice as other assignments, but they are closed.

Lecturer: You give an assignment where the one who formulates the questions knows very well beforehand what the answer is. And then you, as a student, just try to guess the right answer. [...]

In addition, it seems that lecturers that are critical to the possibilities to use MCQs to prevent surface learning do so as part of a group culture, presumably at the faculty level.

Lecturer: This is not the kind of things we want the students to spend their time doing throughout their studies.

Another concern related to the issue of surface learning is that the students would get confused when reading the wrong answers provided on the MCQ test and would memorize them concurrently with the right answers.

Lecturer: An inconvenient which I think few people think about is that when people [students] work with multiple-choice tests, give an answer and later on get to hear what the right answer was, if they have first seen the wrong ones, it has a negative effect in itself, because those [wrong answers] get stored in the same place as the right ones. And then you can get interferences [...].

Several of the interviewed lecturers express a high level of resistance to the use of MCQs, because they are concerned about their propensity to only support surface learning. From the interviews it appears that the lecturers use the tool "as is" for summative assessment and do not consider the possibility to transform it into a tool for formative assessment.

Lecturer: But it is a question of principles. I am not a fan of those multiple-choice questions, and I call them... what is it called? Answer-dropping... that is, a method of elimination, like. (...) I am more a supporter of other types of assignments then, where they [the students] have to express themselves and think a bit. [*In this context, the Norwegian word refers to the process of leaving out answers that are obviously unsuitable, and therefore ending up with a smaller set of possible right answers]*

From the explanations provided from the lecturers it appears that their pedagogical beliefs regarding the affordances of the tool are somewhat preset. They do not seem to envisage other areas of application than surface learning.

Interviewer: Then I would ask you: multiple-choice tool, have you used it [...]?

Lecturer: No, I haven't felt the need for it [...]. In other words, I think that ... it is not my style in a way. I don't teach that way. [...] I have a strong conviction that [the VLE] has nothing to do with ... [...] . My teaching just doesn't fit into a VLE logic.

Those lecturers that use the MCQ tool express clearly their appreciation of the efficiency of such tools to provide a certain type of learning outcomes, i.e. drilling and rote learning.

Lecturer: The point is to get them to read.

Lecturer: At the same time they get [the opportunity to] drill [the material], which they would not have had if they just sat and read the books.

The third design category in Draper's model is **confidence-based marking**, where students are asked to indicate, for each question, their degree of confidence in the answer they give. Many of the lecturers interviewed for this study do not seem to have considered the possibility of implementing techniques enabling confidence-based marking. However, several of them voice their concern about getting false results with MCQ tests, due to students "guessing their way through" the tests.

Lecturer: You might have [a situation where students] are way above the pass line, in spite of having just guessed [the answers] in a rather random manner and they will get a pass grade in the end.

Lecturer: The question is then how to best get prepared for such a - shall we say "guessing competition".

It has to be noted that the actual VLE that was under investigation does not include any tool that would enable confidence-based marking. One lecturer seems to be aware of the issue, and expresses his eagerness to penalize "guessing competitions". He mentions that he has found a roundabout solution to the problem that consists of attributing negative scores to wrong answers, and thereby hoping to reduce the number of guessed answers.

Lecturer: The evaluation system is such that a wrong answer gets penalized. If you only give them [the students] zero points [for a wrong answer], then they have nothing to lose when taking a chance. But if, on the contrary, you say that a wrong answer results in negative points, then if you take a chance, you would have in a way not lost one, but two points.

Although **brain teasers** can function at both individual and group level according to Draper's categorization, we have not found any examples of MCQs used to prompt peer discussion in our data set. However, in one instance, we have found that MCQs have been used as brain teasers in terms of shedding light onto what the students do not know and need to study in more depth.

Lecturer: I think [the multiple choice tools] are nice. [...] the way they are functioning, they look mostly like a nice way to drill. You can work through them and then it makes you understand that there are large amounts of material you don't know.

Draper's fifth category is about **creating MCQs for the purpose of peer discussion**. At one of the faculties of the examined institution MCQs have been used first experimentally and later on more widely as a way of training students. In that faculty, the students are required to use the MCQ function of the VLE to create tests for each other. An example of such a multiple-choice assignment is given by one of the interviewees:

Lecturer: So then they bring in question from anatomy, but also questions that relate to the [academic discipline of] nursing. How should one do this, should one... a patient who has breathing problems, for

example, should he lie on his side, should he stand up, should he sit up, and so on. Or how long is the intestine, is it 3 meters, is it 5 meters, is it 10 meters? Which nutrients are absorbed where...

Originally this approach to teaching was introduced in order to solve a recurring problem that had been identified by several faculty members: the students seemed generally little interested in delving into a curriculum that was considerable in size and not particularly exciting.

Lecturer: They [the students] publish the questions [online] and then they get feedback from their fellow students who are supposed to answer the questions about how relevant and how valid they are regarding the curriculum.

The introduction of a VLE-based MCQ assignment was in a way a motivation trick from the part of the lecturers to increase student engagement and attainment. The experiment was deemed successful, as the MCQ assignment appeared to be both enjoyable and “fun” for the students.

In this example, we see that Drapers’ fifth category has a direct relevance to the pedagogical practice of the interviewees. It is apparent that the students use time to discuss the processes of construction of the multiple-choice questions. In our example, the lecturers go one step further and use the discussion as a basis for deeper understanding of the curriculum. In trying to identify what is important and what is secondary in the presented material, the students achieve a more thorough understanding of the curriculum and of their discipline of study.

It is interesting to note that the processes around MCQ assessment have reached such a high status in the faculty that formal classroom teaching has been reduced in terms of hours in order to liberate time for group discussion around the formulation of MCQs. This reflects the lecturers’ faith in the capacity of MCQs to foster higher order learning and more targeted teaching.

Interviewer: So those comments, do they say... do they come into a portfolio?

Lecturer: [...] Yes, it ends up in a kind of knowledge base where they write their thoughts and ideas about the process of acquiring knowledge through writing, or formulating such a test.

Interviewer: But what is your role in all that? Is it so that you look at the questions before showing them to the students or...?

Lecturer: No. [...] I open [the VLE] and look at the questions, but I don’t necessarily give feedback on the questions, because they [the students] already get so much of that from their fellow students. And since this has to do with certain basic needs related to what they are working with at that very moment, the feedback they get from their fellow students is better, almost, than what I could give to them.

Drapers’ sixth category consists of **having students create MCQ for summative assessment**. We have found at least one formal application of the MCQ that fits into that category. For some of the courses, students are required to develop examination questions. As one of the interviewees mentions, when referring to what he tells his students’ about how they should prepare for final exams:

Lecturer: ... this is supposed to be a voluntary effort in a sense, but there are a lot of good pedagogical reasons for you [the students] to make [the examination questions] yourself, and you learn a lot in doing so, and the whole curriculum becomes part of the agenda at the early stages of the academic year.

He compares this pedagogical method to a much less successful “traditional” approach to examinations where students tend to procrastinate revising for the final exam until the very last minute.

Lecturer: ...instead of what used to happen before, that suddenly someone discovers some book or other at the beginning of June and unfortunately, it is sold out.

Discussion and conclusion

The categorization proposed by Draper could appear to be a strong argument for using MCQ assessment in higher education. However, our study reveals that many of the categories suggested do not find any resonance among our interviewees. We observed that the interviewees were generally little interested in using MCQs, and, in several instances where MCQ assessment actually was used, the focus was often on measuring exact knowledge.

MCQs are used to promote and test traditional skills in a variety of subject areas. However, in most of the cases examined in this study, the assessment style is dominated by a focus on exact knowledge, which appears to encourage surface learning. The data is not giving any indication that lecturers have at all considered using MCQs to attain those higher-order skills that are supposed to emerge when implementing techniques within Draper's first four categories. One explanation could be that the tool used at the institution under examination does not lend itself easily to practices involving self-evaluation based on reasoning and confidence-based marking. It could be tempting to attribute the lack of interest for the MCQs among the lecturers to this limitation of the tool, but our impression is that the reasons for the resistance are more complex and multi-faceted. We have gathered evidence throughout our data set that those lecturers that resist MCQs do so without having spent much time investigating the affordances of the tool.

When it comes to students using MCQs either for peer discussion or as a part of the summative assessment, we experience that our data set is in line with Draper's 5th and 6th categories. The interviewees report practices that are based on peer assessment and formative assessment philosophy. They also report that they aim at and achieve higher-order learning with MCQ assessment. In such instances we observe that the students are thrown into learning situations that are meant to further "new millennium skills" that the lecturers believe to be of paramount importance.

Although all the faculties represented in this study mention the acquisition of digital literacy as part of their strategy plan, the MCQ tool does not seem to represent a significant entry into the world of digital literacy. In spite of the fact that the tool is computer-based, its use is rarely associated with the acquisition of general digital literacy. Its purpose is mainly to get students to acquire knowledge and skills relevant to the curriculum of their study. It is not surprising that study programs preparing for professions such as engineering and nursing do not value highly the knowledge of how MCQ tools function, since those tools are rarely a part of the students' future professional practice. What is more surprising is that the lecturers at the Faculty of Education show little interest in training students to understand the intricacies of the MCQ tool despite the fact that such a tool is an inherent part of the school system at all levels.

The categorization proposed by Draper could be guidance for good practice among lecturers. Nevertheless, our study has revealed that many of the interviewees do not seem to relate to the ideas presented in the model. Further research may be based on new interviews where lecturers would be presented with Draper's categorization and asked to reflect on how the categories may relate to their pedagogical beliefs and which relevance they may have to their pedagogical practice.

Acknowledgement: This paper is based on research of the GOLEM project (Generating learning using an Online Learning Environment as a Medium) and has been made possible by a grant obtained by Norway Open Universities.

References

- Bennett, R. E., Jenkins, F., Persky, H., & Weiss, A. (Writer) (2003). Assessing complex problem solving performances [Article], *Assessment in Education: Principles, Policy & Practice*: Routledge.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working Inside the Black Box: Assessment for Learning in the Classroom*. London: NferNelson.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Maidenhead: Open University Press.
- Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7.
- Draper, S. W. (2009a). Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology*, 40(2), 285-293.
- Draper, S. W. (2009b). What are learners actually regulating when given feedback? *British Journal of Educational Technology*, 40(2), 306-315.

- Erstad, O. (2009). Changing assessment practices and the role of IT. In J. Voogt & G. Knezek (Eds.), *International Handbook of Information Technology in Primary and Secondary Education* (pp. 181-194): Springer Publishing Company.
- Gibbs, G. (2006). Why assessment is changing. In C. Bryan & K. Glegg (Eds.), *Innovative Assessment in Higher Education*. London and New York: Routledge.
- Heinze, A., & Heinze, B. (2009). Blended e-learning skeleton of conversation: Improving formative assessment in undergraduate dissertation supervision. *British Journal of Educational Technology*, 40(2), 294-305.
- Howe, C., McWilliam, D., & Cross, G. (2005). Chance favours only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology*, 96(1), 67-93.
- Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology*, 40(2), 371-385.
- Landauer, T. K., Laham, D., & Foltz, P. (Writer) (2003). Automatic essay assessment [Article], *Assessment in Education: Principles, Policy & Practice*: Routledge.
- Loddington, S., Pond, K., Wilkinson, N., & Willmot, P. (2009). A case study of the development of WebPA: An online peer-moderated marking tool. *British Journal of Educational Technology*, 2, 329-341.
- Mazur, E. (1997). *Peer instruction: a user's manual*. London: Prentice Hall.
- McFarlane, A. (2003). Assessment for the digital age. *Assessment in Education: Principles, Policy & Practice*, 10, 261-266.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed. ed.). Thousand Oaks, California: Sage.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, California: Sage.
- Quellmalz, E. S., & Kozma, R. (Writer) (2003). Designing assessment of learning with technology [Article], *Assessment in Education: Principles, Policy & Practice*: Routledge.
- Ridgway, J., & McCusker, S. (Writer) (2003). Using computers to assess new educational goals [Article], *Assessment in Education: Principles, Policy & Practice*: Routledge.
- Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature Review of E-assessment*. Bristol, England: Futurelab.
- Russell, M., Goldberg, A., & O'Connor, K. (Writer) (2003). Computer-based testing and validity: a look back into the future [Article], *Assessment in Education: Principles, Policy & Practice*: Routledge.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77.
- Shephard, K. (2009). e is for exploration: Assessing hard-to-measure learning outcomes. *British Journal of Educational Technology*, 40(2), 386-398.
- Sieber, V. (2009). Diagnostic online assessment of basic IT skills in 1st-year undergraduates in the Medical Sciences Division, University of Oxford. *British Journal of Educational Technology*, 40(2), 215-226.
- Wirth, J., & Klieme, E. (Writer) (2003). Computer-based assessment of problem solving competence [Article], *Assessment in Education: Principles, Policy & Practice*: Routledge.