An analysis of the impact of an authentic assessment strategy on student performance in a technology-mediated constructivist classroom:
A study revisited

Andrew Scholtz
University of Limpopo, South Africa

ABSTRACT
Assessment performs a number of important and well documented roles in learning environments where it is used as both a formative and a summative tool. However, one of the most contentious roles that assessment plays is its role in high stakes accountability testing. Over the years a degree of standardisation of summative assessment has occurred that appears to satisfy society’s need for certainty about the validity and reliability of summative assessment practices, particularly in the case of high stakes accountability testing. Promotion of competent learners at schools and tertiary institutions depends on the outcome of this assessment, as does the process of warranting learning, while employers rely on these outcomes when deciding on whom to employ. This form of assessment practice has strong roots in the behaviourist paradigm and relies on ‘scientific measurement of ability and achievement’ for its authority. So strong is the hold of the behaviourist approach on summative assessment practices that it is ‘presumed to hold the high ground’ even in constructivist classrooms.

In this paper a study undertaken in 2002 that considered the implementation of a computer-mediated, constructivist learning environment is revisited in light of tensions concerning validity and reliability between the behaviourist-informed measurement community and the authentic assessment practices of the social constructivist community. The results of student performance in the assessment that took place in the original study are reassessed and discussed in terms of the behaviourist versus constructivist debate with respect to assessment. Apart from the obvious wider implications, this debate has particular relevance with respect to institutional online learning implementation via staff development programmes.

Keywords: Assessment; authentic assessment; accountability; validity and reliability; measurement community; constructivist learning environments

INTRODUCTION
This study revisits an assessment strategy employed in a study undertaken in 2002 (Scholtz, 2005) which documented the design, development and implementation of a computer-mediated constructivist learning environment and its effect on students at an historically black institution. Of particular interest to the author is the tension that exists between social constructivist-informed authentic assessment practices and the belief systems and expectations of educators, administrators, employers and parents (Shepard, 2000a: 1; Shepard, 2000b: 6), which justify the continuation of the status quo, supported as it is by the practices of the measurement community (Shepard & Bliem, 1995: 1).

It is important to point out early on in this discussion that the design of the module presented in the original study – and by implication the assessment approach followed – was informed by Herrington & Oliver’s (2000) work on technology-mediated authentic learning environments. Herrington & Oliver’s (2000) work is, in turn, a synthesis of the ideas a number of authors in the social constructivist school, in particular Brown, Collins & Duguid’s (1989) notion of situated
learning and cognitive apprenticeships and Lave & Wenger’s (1991) notion of legitimate peripheral participation within communities of practice.

Social constructivists are adherents to Vygotsky’s Social Development Theory and Blumer’s symbolic interactionist point of view (Kanuka & Anderson, 1998: 60; Kanuka & Anderson, 1999: online). They emphasise the importance of the role of language and communities or groups, with common interests or ‘shared practices’, in the construction of knowledge through interaction (Kanuka & Anderson, 1998: 60; Kanuka & Anderson, 1999: online). In other words, as Kanuka & Anderson (1999: online) point out:

...knowledge is constructed in the context of the environment in which it is encountered through a social and collaborative process using language.

Obviously the theoretical foundation on which this module was developed is important, however the discussion that this paper seeks to stimulate focuses on the issues raised by Shepard in 1991 when she asks why it is that the behaviourist-underpinned approach to assessment of the measurement community is ‘presumed to have the high ground’ (Shepard, 1991: 9).

THEORETICAL FRAMEWORK

The influence of behaviourist psychology on education has endured for more than five decades. While there is evidence that the influence of social constructivism on education practice in the classroom is on the increase, there is also evidence that this influence does not extend to assessment (Shepard, 2000a: 4). On the contrary, Shepard (1991: 1) contends that the implicit beliefs and theories of teachers, administrators and other key role-players, including parents, are so influenced by the dominant paradigm of their formative professional and lived experiences that the contemplation of alternatives to the behaviourist concept of ‘scientific measurement of ability and achievement’ (Shepard, 2000b: 5) is difficult (Shepard & Bliem, 1995: 1).

This is particularly true of high stakes accountability testing, where the results of the assessment determine whether learners are promoted or their learning can be warranted (Knight, 2002: 276). Born out of the need to address ‘embarrassing inconsistencies in teachers’ grading practices’ (Shepard, 2000b: 14), it is the very notions of evidence and fairness that go to the heart of the issue, namely that the behaviourist approach to assessment is ‘presumed to have the high ground’ (Shepard, 1991: 9). Such assumptions shape ‘beliefs about the nature of evidence and principles of fairness’ (Shepard, 2000b: 17).

Behaviourists have, over several decades, developed an approach to testing that they believe measures the ability of learners objectively against a set of norms or criteria designed specifically for that purpose. This approach is based on the classic behaviourist assumption espoused by Skinner that discipline-specific knowledge can be deconstructed into discrete, ‘tightly specified behaviourally-stated objectives’ ((Entwistle, 1988: 8; Shepard, 2000b: 9), the mastery of which must be demonstrated through explicit testing before learners can proceed to the next level. In this way behaviourists applied Thorndike’s principles of scientific measurement (see Thorndike, 1904 and Thorndike, 1927) to these tests in order to standardise their outcomes. This process of ‘making the study of education more scientific’ (Shepard, 2000b: 14) resulted in an increasing confidence in the outcome of the assessment process in the minds of teachers, parents, administrators and politicians alike.

Critics of the behaviourist approach to testing and assessment argue that such tests have had the effect of sustaining the gap between knowing and doing, and the decontextualisation of learning (Brown, Collins & Duguid, 1989: online; Ramsden, 1992: 39; Laurillard, 1993: 15-17; Kings, 1994: online; Herrington & Oliver, 2000: online; Herrington, Reeves, Oliver & Woo, 2004: 4).
Furthermore it is asserted that behaviourist ‘commoditization of learning’ promotes ‘conflicts between learning to know and learning to display knowledge for evaluation’ (Lave & Wenger, 1991: 112). This has, in the opinion of Shepard (2000b: 3), led to the moulding of classroom activities around both the ‘content and format of external standardised tests’, resulting in the ‘complexity and demands of the curriculum’ being lowered and a reduction in the ‘credibility of test scores’.

The social constructivist alternative to behaviourist pedagogy sees learning as the construction of knowledge within the context of real life situations and assessment integrated into the process of learning (Wild & Quinn, 1998: 76-77; Brown, Collins & Duguid, 1989: online; Cognition and Technology Group at Vanderbilt, Learning Technology Center, 1993: 75; Laurillard, 1993: 15; Herrington & Oliver, 2000: online; Shepard, 2000b: 1). In other words, if assessment is to be meaningful it should in some way reflect the practice of the profession, vocation or practice being assessed, while at the same time giving learners the opportunity to demonstrate their knowledge and skills.

Shepard describes this approach to assessment as performance based (Shepard, 2000b: 43), in which:

Teachers’ close assessment of students’ understandings, feedback from peers, and student self-assessment are a part of the social processes that mediate the development of intellectual abilities, construction of knowledge, and formation of students’ identities.

The study that this paper revisits involved the design, development and implementation of an authentic learning environment – and by implication an authentic assessment strategy – based on Herrington & Oliver’s (2000: online) nine characteristics of authentic learning environments, namely that authentic leaning environments should:

1. Provide authentic contexts that reflect the way knowledge will be used in real life;
2. Provide authentic activities;
3. Provide access to expert performances and the modelling of processes;
4. Provide multiple roles and perspectives;
5. Support collaborative construction of knowledge;
6. Provide reflection to enable abstraction to be formed;
7. Provide articulation to enable tacit knowledge to be made explicit;
8. Provide coaching and scaffolding by the teacher at critical times; and,
9. Provide for authentic assessment of learning within the tasks.

The issue under consideration is whether assessment based on social constructivist principles can overcome the concerns of teachers, parents, administrators, politicians and other commentators whose thinking is so influenced by the notion of validity and reliability that is inherent in behaviourist psychology’s concept of ‘scientific measurement of ability and achievement’ (Shepard & Bliem, 1995: 1; Shepard, 2000a: 1; Shepard, 2000b: 6).

THE STUDY REVISITED

One of the questions posed in the original study (Scholtz, 2005) concerned the effect of an authentic assessment strategy in a technology-mediated, constructivist-informed learning environment on the performance of students who participated in this study. When posing this
question one is immediately aware of the tensions between constructivism and behaviourism in this study. Before examining these tensions more thoroughly it is important to briefly describe the module designed for the original study.

The Module

Support for the design of the module that was developed for this study was drawn from a number of theoretical perspectives and represents an attempt to develop a technology-mediated authentic learning environment based on the ideas of Herrington & Oliver (2000), whose work is influenced by both Brown, Collins & Duguid’s (1989) notion cognitive apprenticeship and Lave & Wenger’s (1991) notion of legitimate peripheral participation in communities of practice. The design process also acknowledged the importance of:

- interaction in learning environments as an influence on student attitudes and student achievement (Hillman, Willis & Gunawardena, 1994; Sutton, 2001; and Moore, 1989). The use of online learning environments to promote interaction between learners and content, learners and learners, learners and teachers, learners and the interface is usually intended to satisfy the learner’s need for support (Ally, 2004);

- communication in support of these interactions (Anderson, 2002);

- assessment as central to the learning experience (Brown, et al., 1994; Kings, 1994; Hodgman, 1997 and Rovai, 2000), and its influence on the ‘choice’ of learning made by the learner (Hodgman, 1997; Marton & Säljö, 1984; Dahgren, 1984 and Entwistle, 1988); and,

- the generic outcomes required by the National Qualifications Framework of the South African Qualifications Authority (undated).

At the beginning of the course a group of final year Physiology students were asked to form groups of six. No criteria were used in this process and students were able to choose their group mates as they saw fit. However, the class was informed that participation in the module required a degree of computer literacy and they were advised to ensure that at least one group member was reasonably computer literate, if possible. Each group member was assigned a role within the group by consensus amongst the group members. No particular thought was given to structuring the groups or the roles within the groups other than the generally-acknowledged importance of group work in social constructivist learning environments. Students performing the same function within the group were brought together to learn about their particular function within the group and what was expected of them. Table 1 lists the required roles and concomitant responsibilities.

After dealing with the roles and responsibilities of individuals within a group, the groups were introduced to the tasks they were to undertake. Each task was tackled by two groups so that the students could participate in a peer assessment process with some exposure to the subject matter and a degree of understanding of the topic. In designing the tasks an attempt was made to present these tasks in an authentic a manner as possible, situated in the real world context that physiologists might have to contend with in their working environment.
Table 1: Individual Roles and Responsibilities within a Group

<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Leader</td>
<td>Group leaders were responsible for co-ordinating the group’s activities and the development and implementation of an action plan, in conjunction with group members, in order to ensure that the tasks set were accomplished.</td>
</tr>
<tr>
<td>Internet Researcher</td>
<td>Internet researchers were given a short course on the use of the Internet and pointed to a number of online resources dealing with Internet searches.</td>
</tr>
<tr>
<td>Library Researcher</td>
<td>Library researchers were given a tour of the university library by a subject librarian and were briefed on how to make use of the library to find suitable information.</td>
</tr>
<tr>
<td>Scribe</td>
<td>The scribes were given a short course on the use of MS Word and pointed to a number of online resources that they would find useful in completing their role in the team.</td>
</tr>
<tr>
<td>Presenter</td>
<td>The presenters were given a short course on the use of MS PowerPoint and pointed to a number of resources that they would find useful in completing their role in the team.</td>
</tr>
<tr>
<td>Assessment Co-ordinator</td>
<td>The assessment co-ordinators were advised of their responsibilities as co-ordinators of the assessment processes and their roles in guiding and understanding the processes required to complete the task. They were given access to a resource that explained assessment to them and the difference between formative and summative assessment. The assessment process was explained to this group and assessment rubrics were given to the assessment co-ordinators as guides to the assessment process.</td>
</tr>
</tbody>
</table>

Assessment Strategy

Groups were expected to make use of the Internet and the university library in order to access the resources necessary to successfully complete the task. Each group was expected to submit electronically a five-page typed report on their task, in the format required, which stressed the importance of citations in the text and references at the end of the document. The documents submitted were made available to the class on the module website. These initial submissions became the focus of a formative assessment exercise undertaken by the groups and by a panel of experts made up of the class lecturers, three graduate assistants and the author as facilitator of the module. Each group was required to comment on the submission of the group doing the same task as they were, i.e. peer group assessment. An assessment rubric was made available electronically for the purpose and was completed by groups and the panel of experts alike. This rubric also contained an area where groups could post detailed comments about the submission that they were assessing. Groups were obliged to provide a detailed report justifying their criticisms as well as pointing out where improvements could be made.

In order to ensure that the process of formative assessment undertaken by the peer group was taken seriously the group was assigned a mark for their efforts. These marks were given equally to group members and assigned to a category called ‘Contribution to discussion and assessment of tasks’.

On completion of the formative assessment process, groups were given an opportunity to reflect on the input of their peers and of the subject experts and to reconsider their submission based on what they had learned from both the formative assessment process. This reflective process
culminated in the resubmission of the tasks by the groups. This resubmission was for summative evaluation, which was undertaken by the module lecturers. When undertaking this assessment the lecturers concerned themselves not only with the content but also with how the group had dealt with issues arising from the comments received on their submission. Feedback was given by the lecturers to the groups before completion of the next step, the creation of a presentation.

Subsequently, groups were required to create an oral presentation on their task for delivery to the class. The class and the panel of experts participated in the assessment of the presentation making use of an online rubric designed to guide the assessment process. Participation by the class in this process was assessed and marks allocated to the category ‘Contribution to discussion and assessment of tasks’.

Finally, in order to ensure that students were rewarded for their participation within the group, student-participants were required to assess the contribution of each of their peers within their group. Students could earn or lose up to 12% of the final mark awarded to the group, based on the results of this peer assessment. Students who did not participate in this process were penalised. Students who did not take the process seriously, for example by awarding the same rating to each question contained in the poll or the same rating to all participants in the group, were also penalised, and their ratings were discounted in the final calculation. This was reflected in the assessment category called ‘On-going assessment of attitudes to the module’.

RESULTS

Student performance in the study module, which I will refer to as Module 1, was revisited and compared to student performance in the module following the study module, which I will refer to as Module 2, in order to ascertain whether student participation in a technology-mediated constructivist learning environment had any influence on their performance when compared to performance of the same group of students in a traditionally-presented chalk-and-talk classroom. An exploratory analysis of student performance in these modules using MSExcel indicated that there was a difference in student performance between modules and that the degree to which student performance differed was not uniform throughout the class.

Indeed, the difference in performance between the modules for the class as a whole and the performance of students at the top of the class as determined by their performance in Module 2 was not the same as that of students in the middle of the class or at the bottom of the class. While there are a number of factors that could have been instrumental in the cause of this manifestation, the pattern was compelling enough to warrant further investigation given the tensions between constructivist learning environments and summative assessment practices.

In order to do so the class was divided into tertiles based on their individual performances in Module 2, the follow-on module. A paired samples t-test was undertaken on the performances of the class as a whole in both modules and on the performances of each of the tertiles in both modules using SPSS. The results of this test are given in Table 2.
Table 2: Results of the Paired Samples t-test

<table>
<thead>
<tr>
<th>Tertile</th>
<th>Module</th>
<th>Paired Differences</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Module 1 - Module 2</td>
<td>Mean</td>
<td>38.53</td>
<td>10.06</td>
<td>2.31</td>
<td>33.68</td>
<td>43.37</td>
<td>16.70</td>
</tr>
<tr>
<td>Middle</td>
<td>Module 1 - Module 2</td>
<td>Mean</td>
<td>24.65</td>
<td>7.11</td>
<td>1.59</td>
<td>21.32</td>
<td>27.98</td>
<td>15.50</td>
</tr>
<tr>
<td>High</td>
<td>Module 1 - Module 2</td>
<td>Mean</td>
<td>11.83</td>
<td>7.96</td>
<td>2.30</td>
<td>6.77</td>
<td>16.89</td>
<td>5.15</td>
</tr>
<tr>
<td>All</td>
<td>Module 1 - Module 2</td>
<td>Mean</td>
<td>26.80</td>
<td>13.32</td>
<td>1.87</td>
<td>23.06</td>
<td>30.55</td>
<td>14.37</td>
</tr>
</tbody>
</table>

The paired-samples t-test compares the means of two variables that represent the same group at different times. In this case the two variables are the different approaches taken in the modules in which the group participated, i.e. in Module 1, the study module, students participated in a computer-mediated constructivist classroom, while in Module 2 students participated in a traditionally-presented chalk-and-talk classroom.

Like z-scores, the paired-samples t-test standardises individual items in a population distribution by taking into account the mean and standard deviation of that population, thus allowing for comparisons to be made. From Table 2 the fact that the significance values for the difference between means of each tertile is zero, (i.e. p = 0.00), and the fact that the upper and lower 95% confidence interval do not contain a 0, indicates a significant difference between the means of student performance in each tertile. This also applies to the analysis for the module as a whole, i.e. that there is a significant difference between student performance in each of the modules.

Furthermore, the difference between the means for the performance in each module of the class as a whole is 26.80, while for the students in the middle tertile this difference is 24.65, which is little different from the class as a whole. However, when considering the difference between the means for the students who fell into the low tertile, we see that there is a greater difference between the difference in means between the performance of the class as a whole (26.80) and the difference in means between the performances of students in this tertile (38.53).

The results of this test appear to indicate that students in the low tertile were advantaged by the approach taken in the study module (Module 1) over the approach taken in the follow-on module (Module 2). Finally, when considering what happened to students in the high tertile, we find that the difference in their performance (11.83) when compared to the difference in means between the modules as a whole (26.80) was a great deal smaller than for the difference in means between the modules.

The results of this test appear to indicate that students in the top tertile were disadvantaged by the approach taken and did not, or were not able to fulfil their potential in the study module (Module 1) when compared to their performance in the follow-on module (Module 2). These results of the paired samples t-test analysis would seem to indicate that the group approach
Student performance in a technology-mediated constructivist classroom

taken in Module 1 seems to have a 'uniforming' effect on student performance when compared to student performance in a traditional chalk-and-talk classroom.

A One Way Anova analysis of the means was performed on each of the tertiles within each module in an attempt to confirm this pattern. The results are given in Table 3.

Table 3: One Way Anova Analysis of Means

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>Between Groups</td>
<td>113.493</td>
<td>2</td>
<td>56.747</td>
<td>1.629</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>1671.801</td>
<td>48</td>
<td>34.829</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1785.294</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 2</td>
<td>Between Groups</td>
<td>7037.830</td>
<td>2</td>
<td>3518.915</td>
<td>84.490</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>1999.151</td>
<td>48</td>
<td>41.649</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>9036.980</td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the Anova analysis of the modules one can see that the difference in means between the tertiles in Module 1 was not significant (p<0.05). While for Module 2 the difference in means between the tertiles was indeed significant (p = 0). This indicates that students in the bottom tertile performed statistically significantly worse than those in the middle tertile. Students in the top tertile performed significantly better than those in the middle tertile. In other words, there is a significant difference in student performance depending on which tertile students find themselves. When the situation in Module 1 is considered there is no statistically significant difference between the mean results obtained by the students in any of the tertiles.

This would further suggest that the assessment strategy in Module 1 had a the effect of advantaging the poorer performing student as determined by student performance in Module 2; had little effect on the participants in the middle tertile and disadvantaged the top students as determined by student performance in Module 2. This seems to be a further indication of the 'uniforming' effect on student performance of the group approach taken in Module 1 when compared to student performance in a traditional chalk-and-talk classroom.

DISCUSSION

The assessment approach used in the study can certainly be criticised on a number of counts. Firstly, the strong reliance on group assessment used needs to be reconsidered to provide students with opportunities to show individually what they are capable of doing. Secondly, this preoccupation with group assessment will tend to have a 'uniforming' effect on the performance of a group and, ultimately, on the performance of a class. It is conceivable that the statistical results obtained may have been determined by the low limit of 12% which was set for the maximum variation between the group mark and the individual mark. Finally, it is clear that more consideration needs to be given to the theory with respect to authentic tasks and collaboration in
authentic learning environments.

However, these criticisms should not detract from the issue at hand, namely that:

‘The dominance of objective tests has . . . shaped beliefs about the nature of evidence and principles of fairness’ (Shepard, 2000b: 17).

Clearly, the results obtained from revisiting aspects of this earlier study – no matter how flawed they might be – lend credence to the concerns that the measurement community have about authentic assessment practices, particularly with respect to the validity and reliability of high stakes summative assessment practices. It would appear that these concerns regarding assessment are shared by many who otherwise embrace social constructivist learning environments, hence the concern raised by Shepard (2000: 5) and others that traditional testing remains the predominant form of assessment, even in constructivist classrooms. This is of particular concern given that the literature is fairly unanimous in its support of social constructivism as the pedagogy of choice in support of technology-mediation in learning.

Successfully challenging the implicit beliefs and theories of teachers, administrators and other key role-players is therefore a vital step if alternative or authentic assessment practices are to gain acceptance in the modern classroom. In order to do so analysis of these assessment practices need to present a more convincing picture, particularly as far as the validity and reliability of the outcome of these practices are concerned. It is interesting that, while constructivist literature is fairly clear about what learning is and the sort of learning environments we need to create in order to bring learning about, little seems to be written about how we determine whether learning is, in fact, taking place and, if so, to what degree.

If authentic assessment is to acquire the sort of legitimacy that the assessment practices of the measurement community have acquired then we as critics of these assessment practices need to find ways and means of confronting the criticisms levelled at alternative assessment.

CONCLUSION

This paper attempted to highlight the sort of concerns that psychometricians have with assessment in constructivist learning environments, particularly with respect to high stakes accountability testing. The results of the analysis undertaken in this study revisited indicate that an argument can be made that stronger students, academically speaking, were disadvantaged by the assessment strategy employed in the study, while weaker students were advantaged. Exponents of alternative assessment strategies are clearly convinced that these strategies more fairly reflect Shepard’s (200b: 17) ‘nature of evidence and principles of fairness’. However, it is this author’s understanding that a great deal more energy needs to go into consideration of the issues surrounding high stakes accountability testing and the implicit beliefs and theories of all participants and stakeholders in that assessment, if alternative assessment practices are to play a meaningful and convincing role in assessment in general, and high stakes accountability assessment in particular.
REFERENCES


Hillman, D. C. A., Willis, D. J. & Gunawardena, C. N. (1994). Learner-Interface Interaction in Distance Education: An Extension of Contemporary Models and Strategies or Practitioners. The American Journal of Distance Education, 8(20), 30-42.


