Qualitative assessment across language barriers: An action research study

Johannes C. Cronjé
Faculty of Informatics and Design, Cape Peninsula University of Technology, PO Box 652, Cape Town, South Africa, 8000 // Tel: +27 21 460 3877 // Fax +27 21 460 3553 // johannes.cronje@gmail.com

ABSTRACT
If students cannot express themselves in the language of the assessor, and if the assessor is not familiar with the cultural constraints within which students operate, it is difficult for the assessor to collect evidence of adequate performance. This article describes the assessment of three digital artefacts where the assessor strove to understand, rather than rate the students and their achievement. Three distinct approaches were followed. The first was positivist, rubric-based. The second was post-structural/interpretive, with groupwork and peer assessment. The third was feminist/holistic relying principally on students’ own presentation of video material of what they did. Data sources include artifacts produced by students, assessment rubrics and assessment scores. Results indicate that qualitative assessment with criteria negotiated between the instructor/assessor, the twelve learners and their peers over a period of two years, allowed for a ‘thick description’ of the evaluation experience that adds to student development. Further research needs to be conducted into standardizing the procedures of auditing assessments, rather than standardizing procedures of assessing.

Keywords
Language barriers; qualitative assessment; blended learning; computers in education

Introduction and background
This article is a sequel to the story of how we taught a two-year Master’s programme in Information and Communication Technology (ICT) for Education in Khartoum, Sudan. (Cronje, 2006). It considers the continuous assessment strategies used in three of the courses. The article is built upon Brenda Johnston’s “examination of different approaches to agreement over outcomes” (Johnston, 2004) and is presented here as contribution towards beginning to fill some of the “many gaps in the research on innovative assessment methods such as portfolios” (Johnston, 2004, p. 396). The article also enters in dialogue with Wilkerson & Lang, (2003) who present very useful caveats when using portfolios in “high stakes” assessments – such as for graduation or promotion reasons. The specific application of the research lies in the domain of assessment across cultural and linguistic barriers, for as is pointed out by MacKinnon & Manathunga, “If our assessment continues to be based upon a Western template of knowledge that only values Western ways of knowing and learning, all our lip service to developing interculturally competent students is meaningless” (p. 132).

The purpose of this article is not to “prove” or “reject” any cross-cultural assessment theories. Neither is it an attempt at reporting quantitative, transferable results but rather to share the story of how we dealt with a practical manifestation of the above problem that arose when two colleagues from Pretoria, South Africa, and I presented a Masters’ course in Information and Communication Technology (ICT) for Education in Khartoum, Sudan. A comprehensive account of the instructional design and presentation of the course has been published in the Journal of Educational Technology and Society (Cronje, 2006). One table and two figures from that article are repeated here. In the previous article they were used to explain the instructional design, whereas they are used here to explain the assessment strategies and outcomes.

The key research question addressed in this article is: “How could we assess fairly a group of students with whom we did not share a common first language, or a common culture?” Or more succinctly, how did the instructors and students create common understanding across language and cultural barriers?

Linguistically the problem was that the students could understand us reasonably well. They could read English, and they could, in their groups, support one another in the reading and interpretation of English articles, but they were shy to speak English and their writing was poor. Thus, even if they knew and understood the content, and even if they were able to apply it, their ability to display that insight was marred by their inability to write clear, comprehensible English. Culturally one of the first obstacles was not behavioural, but aesthetic. What they regarded as beautiful in terms of screen design we regarded as gaudy. What we regarded as elegant, they considered bland.
At a deeper level, though, there was also the suspicion that we were foreign colonialists coming to impose our value systems upon them, and that we would use assessment as a “technology” (Delandshere, 2001) to enforce our ways upon them.

The initiative came from the Vice-chancellor of their university who, after a world-wide search, had selected our programme because it was developed in Africa, based on an American model, and delivered in English. Thus, although their institution chose us, they were suspicious of us. We had to establish culturally sensitive interpretations that would lead to a commonality of expectations.

The conceptual model we devised to understand our situation is shown in figure 1.

![Figure 1: The cultural divide](image)

At the top left of the figure is local relevance and application. At the top right is the global situation. The more global something becomes, the harder it may be for locals to see its relevance. The more local things become the less globally competitive or useful they might be. Any emphasis on increasing either local relevance or global competitiveness could increase the divide. Providing too much localization could disempower the community by reducing their international competitiveness. Too much global emphasis may distract from local culture. However, the two sides come together at the bottom. Thus we argued that, in order for us to reach commonality we needed to descend into the canyon to find what is fundamental to what we were doing and work from there. Instead of assessing computer screen design for the “beauty” of colour combinations, for instance, we assessed for adherence to acceptable norms such as allowing for colour-blindness, placing background colours such as blue and green in the background and foreground colours such as red and yellow in the foreground, to avoid such colours from “bleeding” because of the physical construction of the human eye.

It was towards the end of the two-year program that the third member of our teaching team mentioned to me how impressed she was with the standard of the students. She was the last instructor to visit them and had only been brought onto the programme towards the end. Thus she was unaware of any of their previous work, and, in a sense, acted as an external evaluator. She was a regular presenter of the South African version of the course with at least six years’ experience of presenting coursework to Master students in South Africa.

The question then arose, “how can she say that?” They write poorly. They are shy. They don’t speak English well. Yet, all three of us knew that they were producing excellent work; but without written proof, how did we know? What accounted for our gut feeling that these students were equal to or better than their South African counterparts on whom we have based our development work, and who formed the baseline of our comparative gut-feel assessment? Or, more precisely, what evidence did we have to substantiate our claim that these students were masters of ICT in Education?

Three sub-questions drive the “scenes” that will be presented later. The first question is: “How detailed should we make the briefs and assessment criteria for portfolios?” The second is: “What is the value of portfolios that are created by teams?” The third question is: “How can the assessor of a portfolio gain insight into the process by which the portfolio was created?”
**Literature survey**

Delandshere, (2001) points out that often assessment is primarily used as a “technology” to exclude those whom the assessing authority deems unsuitable. She continues that “while there have been important technical developments (e.g. item response theory, item-biases) and technological changes (e.g. computer-adaptive testing) the procedures used have remained fundamentally the same with tests being the primary method of educational assessment” (Delandshere, 2001, p.115). Cross-culturally test bias has been identified as a major contributor to unfair exclusion. This bias is further complicated by the relationship between the individual and the group. Skiba, Knesting & Bush, identify “construct validity, content or item bias, inappropriate standardization samples, examiner and language bias, and differential predictive validity” (2002, p.63) as important sources of test bias. They caution further that: “If test results are treated solely as a representation of inherent individual contribution to aptitude, ignoring cultural and educational factors … test scores may yield inaccurate interpretations, even if the tests themselves can be shown to be unbiased” (2002, p. 75).

For us the challenge lay in assessment that was built upon a common understanding between students and instructors, despite cultural differences, of what would be assessed. In his search for such commonality in assessment at a bilingual (English/French) university in Cameroon, Amin (2002) identified six “factors which students consider important in the evaluation of courses” (p. 281). These factors were identified as Rapport with students; Evaluation of Assignments and feedback; Workload; Availability of Help; Course Organisation; and Language or Communications (Amin, 2002, p. 290).

The diverse nature of qualitative assessment makes it attractive when assessing in a cross-cultural situation, since its use of longitudinal evidence derived from multiple sources provides a rich picture of the teaching and learning reality. It also allows for a multiplicity of learning contexts that are more likely to offer a fair and authentic accounts of the potential and performance of different kinds of students. (Tigelaar, Dolmans, Wolfhagen, & van der Vleuten, 2005, p. 596).

In a richly nuanced environment the problem now becomes one of structure and standardisation so that students and assessors work from a common base. Carraccio, (2004, p386) cautions that unstructured portfolios make it difficult to achieve acceptable standards of reliability and validity in educational measurement. MacKinnon & Manathunga’s (2003) observe that “students all stated that what was required of them needs to be stated clearly and unambiguously” (p. 138). Even more seriously Wikerson & Lang (2003) argue that, should portfolios be used for certification purposes, there could be serious legal implications if the assessors cannot demonstrate that their standards of assessment are comparable to psychometric standards. Here Driessen, van der Vleuten, Schuwirth, van Tartwijk, & Vermunt, (2005, p.216) provide valuable pointers: Their strategies for realising dependability – the pendant of reliability – involve establishing an audit trail (i.e. documentation of the assessment process to enable external checks) and carrying out a dependability audit (i.e. quality assessment procedures with an external auditor).

Standardisation presents its own problems. Johnston, (2004) warns against a tendency by independent evaluators to self-moderate when assessing with a rubric, thus leading to higher inter-rater reliability, but lower validity, while Driessen, et al. claim that “detailed checklists can easily trivialise assessment” (2005, p.215). Thus the question driving the first “scene” asks exactly how clear should the instructions be – should we, for instance supply an assessment rubric together with the project brief? In this way students will know exactly what is expected. But, on the other hand, the assessment may be reduced to an almost mechanical checking off of items on the list.

In our second “scene” we tried to reduce the rigidity of “checklist” style assessment by adding team support. Driessen, et al. (2005, p.216) describe a multi-faceted approach using qualitative methods, where “the following three strategies are important for reaching credibility: triangulation (combining different information sources); prolonged engagement (sufficient time investment by the researcher), and member checking (testing the data with the members of the group from which they were collected).”

The third “scene” approaches another problem with portfolios, which is that while they give a rich description of the product that a student is able to create, they might tell us nothing about the process that the student followed, or about what the student understands of the underlying assumptions. At worst we may have no evidence that the portfolio is indeed the work of the student who submits it. The solution may lie in the way in which a digital portfolio is composed. Instead of having a single focus the portfolio could be based on a number of underlying
theoretical assumptions. Two theoretical perspectives informed this research. The first is an integrative approach to objectivist and constructivist learning (Cronje, 2006) and the second is Johnston’s (2004) four approaches to assessment of portfolios: positivist, post-structuralist, interpretive and feminist.

In an attempt to break the polarisation between objectivist and constructivist learning outcomes Cronje (2006) presents a four quadrant model in which the two main approaches to learning, generative and supplantive, are plotted at right angles, instead of being opposites along a straight line. Four quadrants emerge. The injection quadrant is high in supplantive, objectivist or positivist elements and relies heavily upon direct instruction. The construction quadrant, on the other hand, is high in generative, constructivist activity and learners learn by doing. Learning such as the kind described by Papert (1993), where learners engage in various construction projects in order to learn the underlying mathematics, would occur in this quadrant. The immersion quadrant explains what happens when learners are “thrown into the deep end” without any overt planning, and in the integration quadrant learners are taught directly, and allowed to explore, depending on the nature of the situation.

Followers of positivist approaches to assessment according to Johnston (2004) believe in the possibility of objectivity, scientific measurement and certainty (p 397). They strive for reliability of assessments that are done without any knowledge of the student being assessed, or of the ratings of fellow assessors. This approach closely matches Cronje’s (2006) injection quadrant. For Johnston “A central feature of interpretivist approaches to assessment is that realities, especially social realities, are perceived as mental constructions or interpretations, rather than absolute, objective truths” (Johnston 2004, p.399). This approach corresponds with Cronje’s (2006) construction quadrant.

Post-structuralist approaches, on the other hand “challenge the notion of unified, universal, master narratives, together with their embedded power relationships favouring particular groups” (Johnston 2004, p.398). She continues that “For many post-structuralists, the notion of summative judgements of individuals is so problematic and alien that they tend to focus more on formative, developmental and collaborative assessment” (Johnston, 2004, p. 399). This may match the immersion quadrant. Johnston points out that a feminist assessment system “should strive to be fair and sensitive to these differences and their implications, in order to avoid the systematic advantaging or disadvantaging of particular groups” (2004, p. 400). She points out that the key lies in development rather than accountability. Feminists are more likely to be concerned with an assessment that is sensitive to differences in race, gender class and age “in order to avoid the systematic advantaging or disadvantaging of particular groups” (Johnston, 2004, p. 400). This is likely to be the quadrant of Integration (Cronje 2006).
Methodology

For this research an action research methodology (Cohen, Manion & Morrisson, 2000) was followed in the form of a disciplined self-reflective inquiry into a small scale, real-world intervention in an attempt to reform and improve practice. It is reflective in that it strives to bridge the gap between theory and practice, and is done by a practitioner focusing on his own work, combining diagnosis with reflection and focusing on practical issues. Action research was selected as a methodology because of the position of the researcher as a participant in the process. In this way the research was conducted from an internal, rather than an external perspective. On the down side, of course, this means that the researcher can make no claim of impartiality, and the extent to which the results reported here can be transferred elsewhere is limited. Nevertheless, as has been stated before the purpose of this research is not to “prove” but rather to share, and in this respect it is hoped that the results of this action research study will resonate with results obtained by other researchers.

The three “scenes” presented here are drawn from courses in a Masters’ Degree programme in Information and Communication Technologies (ICT) for Education, presented by a professor from South Africa, to twelve students of the Sudan University of Science and Technology (SUST), Khartoum over a period of two years. Pretoria, South Africa is a historically Christian city with English as its lingua franca. Khartoum, Sudan is a predominantly Muslim city with Arabic as its dominant (if not only) language.

Twelve promising graduates of the education faculty of SUST were selected to follow the program, after having attended six months of intensive English language and computer literacy training. I have been a professor of computers in education since 1994, with my main function being the co-ordination and presentation of a two-year coursework Masters degree with a mini-dissertation. In 1994 I read, and met with Seymour Papert and started working in a “constructionist” fashion. Instead of lectures I would usually give students learning tasks involving some development of technology – from which I hoped that they would construct the knowledge for themselves. In 1996 I took this practice onto the Internet in that many of the work done by my students would be web-based (Cronje 1997). The work we did in Sudan was an extension of this approach.

Assessment strategies varied throughout the course but students were required to construct a digital portfolio containing all work produced during the course. The principal data sources are artefacts produced by students during selected courses of the programme as well as the evaluation instruments and evaluation results produced during the assessment. Interpretation of data was qualitative and involved a close scrutiny to identify trends and draw conclusions.

Findings

The following section will discuss the three evaluation scenes that were selected. The first illustrates what happened when a positivist approach was followed, functioning in the injection quadrant. The second shows the implications of a structuralist/interpretivist approach in the construction quadrant, and the third tends towards a holistic/feminist approach in the integration quadrant. The scenes depict the creation and evaluation of digital artefacts that were included in the portfolios of the students during the action research cycles in which we attempted to gather and refine the evidence that allowed us to report on the progress of the twelve students on the programme. The cyclic nature of action research makes it hard to separate the description of the findings from the actual activities from which those findings were derived. Thus a narrative reporting style will be followed, and the rhythm of each scene will be “we found… so we did… so we found…”

Scene one: A mark out of 200

The first graded assignment given to the students was to conduct a dynamic evaluation of a piece of educational software of their own choice and to comment on its suitability for a given target population. The resultant essay was the main element of the portfolio for a course on “The evaluation of educational software and its effect on learning” the course served partly to introduce students to educational software, but mainly to sharpen their skills in empirical research methodology.
The students were not going to have face-to-face contact with me again for at least eight weeks. They had never handed a project to a stranger before. They were uncertain of their English writing skills. Their resultant high levels of anxiety made it necessary to provide more scaffolding than I would have given my South African students. It was clear that my usual approach of setting a task and letting students learn from their mistakes would be more than they could bear.

Discussions with the class indicated that a positivist approach with very clear boundaries would be more acceptable to them. As a result I decided to do three things. I created an article template that contained typical headings and a few useful phrases, as well as one or two citations in the text and in the bibliography that would serve as examples of how to cite and refer. I then created an assessment rubric in the form of a spreadsheet, and finally divided the students into peer support groups of three.

The assessment rubric contained 54 assessment statements with rating scales ranging from 0-1 to 0-5, depending on the criterion being assessed. For instance, the first statement was “Introduction provides clear program information (0-1)”. The template contained an example stating the name of the program, the author, the publisher and the technical specifications for the computer on which it would run. If the student supplies this information, then one mark is given, else nothing. Item 10 and 11, on the other hand were given a 0 to 5 rating (“Literature review logically arranged in sub-sections” and “Supporting and contradicting literature cited”). Rubric statements covered all aspects of a typical assessment of research papers, such as introduction and problem statement, (seven items, 14 points), literature survey, (six items, 25 points) description and defence of method (17 items, 62 points), reporting of findings (six items, 24 points), quality of interpretation and conclusions (six items, 30 points), technical aspects such as tables, and referencing style (eight items, 25 points) grammar, spelling and general impression, though these were deliberately given a much lower weighting so that students would not be penalized for poor English (four items, 20 points). The rubric added up to 200 points, which could be divided by two to make a percentage. The unusually high number of points awarded to the description and defence of the method is accounted for by the fact that the main aim of the course was to teach evaluation methods, and thus I was more interested in the methods followed than in the findings obtained.

Students were expected to do static evaluations of educational software in groups of three working on one computer and going through the software, rating it according to a checklist. Each student would each select one specific title for a dynamic evaluation on an actual target population. They would discuss the appropriate research methodology in their teams, but conduct the evaluation individually and write the evaluation report individually, using the template. Each student would submit the report to the two other members of the team who would rate it according to the rubric and give feedback. The student could then improve on the essay, self-evaluate it according to the rubric and submit to me the essay, together with the three assessments.

| Table 1: Student self-assessment compared to my assessment (Cronje 2006, p. 281) |
|---------------------------------------------|---------------------------------------------|-----|
| Student | Own grade | My grade | Difference |
| Md | 83 | 86 | -3 |
| Yo | 91 | 86 | 5 |
| Et | 89 | 85 | 4 |
| Ta | 68 | 82 | -14 |
| Om | 70 | 82 | -12 |
| Id | 76 | 76 | 0 |
| Ha | 75 | 75 | 0 |
| Ib | 70 | 73 | -3 |
| Ma | 74 | 69 | 5 |
| Ab | 70 | 68 | 2 |
| Ia | 70 | 67 | 3 |
| Ka | 53 | 51 | 2 |

I then assessed each essay and compared my assessment with that of the student. Should the two sets of grades be within 5% of each other; I would award the student the average of the two. If the student’s grade was too low I
would moderate. If the student’s grade was too high, I would suggest improvements until the grades came within 5%.
The peer grade was disregarded, as the students were able to work on the essay after having seen the peer grading.
The result of the grading can be seen in table 1 (Also reported in Cronje, 2006). The abbreviations relate to the
names of the students.

As it turned out it was not necessary to moderate. The two students who underrated themselves were two of the
better students, and their underrating can be ascribed to a lack of self-confidence, as I determined during interviews
with them afterwards. I believe this to be a product of high uncertainty avoidance, coupled with high power
distance. The students were not sure that I would assign so high a grade and thus allocated themselves a lower grade
which they considered more “likely”. Also they thought it would be arrogant to give themselves such a high mark.

I was initially worried that the grades were too high. In the South African system according to which I worked a pass
is 50%, 75% is a distinction, and 80% is usually as good as it gets. Here all the students passed, two thirds obtained
distinctions, and almost half got more than “as good as it gets”. However, I kept the grades as they were, because I
knew the students had been selected based on their academic prowess. Ka’s very low grade can be ascribed to the
fact that he had been ill and had missed the initial set of lectures and the discussion of the template and the rubric.

What was disconcerting, however, was the ‘mechanical’ nature of the essays. They were all close adaptations of the
template and very clearly written to the rubric. Thus I had removed much creativity from the process and made the
students create “instant” essays that were good, but dull. The assessment could possibly be regarded as “trivial”
(Driessen, et al. 2005). Nevertheless, despite what I considered to be too much “hand holding” already, the students
still emailed me frequently with questions and requests for comment on their pieces as they were working.

Scene two: Spreading the sheets

Based on my disappointment with the essays despite their high grades, I decided to move closer to my usual
problem-based or “constructionist” (Papert, 1993) way of facilitating learning by letting students create a digital
artefact, and letting them learn by doing. Another reason for this approach was that the work to be covered was on
“Tutorials, Drills, Simulations and Games” taken directly from the text book (Alessi & Trollip 2001). I also knew
that letting students construct digital artefacts as the foundation of a learning task would not work well in my
absence, and thus arranged for them to engage in a cooperative learning event designed according to the “Jigsaw”
model of Johnson, Johnson & Holubec (1994). Students were divided into home groups and expert groups. Each
home group had to produce a spreadsheet that would contain a number of statements and a yes/no checkbox. In
response to the checkboxes the spreadsheet would suggest the extent to which a given situation would best be served
either by a tutorial, drill, simulation or game. They also had to design the aesthetic aspects of the spreadsheet, and
program the checkboxes. The expert groups were responsible for determining the criteria for tutorials, drills,
simulations and games respectively. Leaders of the home groups were assigned based on their spreadsheet skills,
while leaders of the expert groups were selected based on their language skills. In this way they could better assist
their team mates in terms of their zones of proximal development.

Once the expert groups had brought together the statements for each checklist the home groups negotiated to
determine the values that the spreadsheet would assign to each category (tutorial, drill, simulation or game). Finally
the home groups worked through one another’s efforts by way of summative and peer evaluation. The final grade
was binary: successfully completed or more work required.

The screenshots of the various sheets, presented as figures 3 to 5 show that, as was the case in the essay with the
rubric and template, the degree of similarity was very high. This could be because of the cooperative learning groups,
and because the outcome was mastery of the same textbook content and spreadsheet skills rather than originality. As
it was a group project the contribution to the final grade was minimal. Students were given a range of other
individual tasks that fall beyond the scope of this article.

The first two spreadsheets (Figure 3 – Also shown in Cronje, 2006) fulfilled all the requirements. They contained 39
and 40 questions respectively, with tick boxes for a yes response. We ran a few scenarios through them, and they
provided useful suggestions that seemed adequate. The two sheets contained essentially the same questions as but in
different sequences as these had been compiled in the expert groups. The weighting of the responses was also
slightly different. The group who called themselves “Arnouba” separated the graph from the numeric information, using a table as well as a graph. The background colour of the spreadsheet was pastel green, but the graph had a bright blue background, while the table with the numeric values had a grey background. The colours of the letters spelling tutorials, drills, simulations, games corresponded to the colours of the bar graph. It is of note that, while the spreadsheet itself is justified to the left, the table and the graph both read from right to left instead of the conventional Western way. This is probably due to the fact that when set to Arabic, MSOffice defaults to right justification and, while it is easy to change the justification of the spreadsheet, it is harder to set the justification of the graphs. The table giving the total scores, however, was done by hand but also from right to left – in the Arabic way.

Figure 3. Methods.xls and Arnouba.xls
The spreadsheets produced by Ka’s team and Yo’s team (Figure 4) used pie charts rather than bar graphs or histograms for the graphic representation. They also chose to reveal the weighting that they assigned to the four modes of computer-based education for each statement.

The spreadsheet of Ma’s group (Figure 5) had a number of problems. Firstly the check boxes were not neatly aligned, the questions were not shuffled, but placed in groups according to the mode, and the pie chart measured...
every item, not giving any result. It was clear that more work was required on the spreadsheet, but their knowledge of tutorials, drills, simulations and games appeared adequate.

<table>
<thead>
<tr>
<th>Drills</th>
<th>Yes/No</th>
<th>Drills</th>
<th>Games</th>
<th>Tutorials</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it useful for learning practice?</td>
<td>Yes</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Does it depend on existing knowledge?</td>
<td>Yes</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Is it necessary to provide feedback for various correct/incorrect responses?</td>
<td>Yes</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Does it use feedback for assessing learner performance?</td>
<td>Yes</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Does it provide a context for learner?</td>
<td>Yes</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Does it depend on determining learning goals?</td>
<td>Yes</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Can it be done in sub-groups?</td>
<td>Yes</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Is there need to store data during the activity?</td>
<td>Yes</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Is it based on creating competition between learners?</td>
<td>Yes</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Is it useful for collaborative learning style?</td>
<td>Yes</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

| Games | | | | | |
|--------|--------|--------|--------|--------|
| Is the subject complex and boring? | Yes | 2 | 5 | 4 | 6 |
| Is the subject relevant to the children? | No | 3 | 0 | 4 | 2 |
| Does the subject motivate learners? | Yes | 4 | 4 | 4 | 6 |
| Is it required for understanding of human nature? | Yes | 4 | 4 | 3 | 6 |
| Does it contain moral information? | Yes | 4 | 4 | 3 | 6 |
| Is the knowledge integrated across content area? | No | 3 | 5 | 4 | 5 |
| Does it has a relationship with business study for mature learners? | Yes | 4 | 4 | 4 | 5 |
| Does it make learners invest more effort and process information? | Yes | 4 | 4 | 3 | 6 |
| Does it contribute knowledge and skills? | Yes | 5 | 2 | 2 | 6 |

| Simulation | | | | | |
|-------------|--------|--------|--------|--------|
| Doing it in real life is unsafe? | Yes | 0 | 2 | 0 | 5 |
| Is it expensive to do in real life? | Yes | 2 | 4 | 0 | 5 |
| Does it teach learner how to do something? | Yes | 1 | 4 | 0 | 5 |
| Does it teach the learner how something works? | Yes | 0 | 3 | 2 | 5 |
| Do you want to increase efficiency? | Yes | 0 | 1 | 2 | 5 |
| Do you want to increase learning transfer? | Yes | 4 | 3 | 2 | 5 |
| Is it complex to do in real life? | Yes | 0 | 2 | 3 | 5 |
| Are there effective distracting in real life? | Yes | 0 | 2 | 3 | 5 |
| Is it long in real life and need simplification? | Yes | 0 | 2 | 3 | 5 |
| Do we want to increase motivation? | Yes | 4 | 4 | 3 | 5 |

The spreadsheet exercise showed how the home groups and expert groups negotiated meaning. The fact that the groups had to know enough to be able to do a given task placed the primary responsibility for the assessment on the learner. Instead of the instructor asking “Is there enough evidence that the students have mastered the content?” the students were asking “Do we have enough knowledge to complete the task?” Interpretive assessment worked well in this instance because the instructor was present to give feedback at all times and the assessment was for formative, rather than summative purposes – there being other tasks for final assessment.

Scene three: At the movies

A course on “Learning theory for computers in education” required students to follow Gagne’s (1987) (essentially objectivist) events of instruction to produce a constructivist learning event. They were then to present the event and report on what they had learnt. Students had to go through the entire “analysis, design, development, implementation and evaluation” cycle. Not only did they have to prepare and present a “lesson”, they also had to prepare computer-based support materials so that the emphasis would be on the learners’ construction of their own meaning.

From previous experience with my South African students I knew that this was a difficult task, easily misinterpreted. Students tended to concentrate too much on the development of the technology and too little on the activities that would be required of their learners. Moreover, despite their best intentions the students ended up indulging in a significant proportion of lecturing. I therefore decided that the analysis and design phase would take place during the contact session with the development and implementation being done while I am away, and the evaluation during my subsequent contact visit.

At all times the students would be involved in peer assessment and cooperative learning, while the final evaluation and assessment would take the form of a 10 minute video presentation by the students explaining to me and to their peers what they did and what they learnt.
We started with the students working in four groups of three, each presenting the background to their chosen field, and engaging in discussion until the lesson theme had been established. We then worked in the whole group of 12, during which students explained the concept of their lessons to each other and to me, and we reached a consensus on the basic outline of the lesson. For homework students then filled out a lesson plan called a “learning event protocol” according to a template that I supplied. They also had some time to design and present a prototype of the spreadsheet, database, slideshow or word processing document that they would use as the technological component of their learning events. These were also discussed and students were given another opportunity to refine them. A deadline was set for the completion of the preparation. Students would email me a complete package consisting of the learning event protocol, the supporting material and the evaluation instrument with which they would finally evaluate their learning events. I supplied feedback on a mailing list, so that every student could read the feedback that I had given to each individual. This was both in the interest of transparency and in the hope that they would learn from the feedback to their peers.

The groups of three were involved in the presentation and recording of the learning events. One facilitated the event. Another acted as an assistant to help with logistical and technological problems and the third would record the entire event on video; thus each student was involved in three presentations in different roles.

Learning Event Protocol
Subject area: Maths, logical functions.
Expected outcomes:

After this learning event, learner should be able to:

1. Know the logical functions "AND " and "OR".
2. Using "AND" and "OR" to linking between two statements.
3. Distinguish between outcome to the arguments when the link is "AND".
4. Distinguish between outcome to the arguments when the link is "OR".
5. Distinguish between symbol of "AND" and symbol of "OR".

<table>
<thead>
<tr>
<th>Time</th>
<th>Learner activity</th>
<th>Facilitator activity</th>
<th>Resource</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 min</td>
<td>Navigation through the lesson.</td>
<td>Encourage as many different types of complex statements.</td>
<td>Slideshow, Mouse, Click to open hyperlink.</td>
<td>Make learners aware that there are different kinds of complex statements.</td>
</tr>
<tr>
<td></td>
<td>Evaluate the statements.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observe the events.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The color of statement box help learner to know the</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. The original learning event protocol

Each student had to write an essay explaining the rationale for the lesson and its design against a background of the literature, and then present their findings, conclusions and recommendations. Once again the main focus of the essay would not be on language, but rather on the quality of their interpretation of what happened. Finally the video was
edited and presented to the class upon my subsequent visit. In this way we could all see firsthand what had occurred as the student took us through step-by-step.

Here follows the discussion of the progress of one student. She was not the best student, nor was she the worst, but the case is typical of what happened in the course.

Figure 6 shows the original learning event protocol presented by the student. After some discussion the first learning outcome “Know the logical functions ‘and’ and ‘or’” was retained, as the group felt that, although “know” could not be measured, the other outcomes explained the term. We also felt that the lesson protocol should break down the time more carefully – a 20-minute interval was too long. In the refined version, however, the student went to the other extreme and tried to plan for every 2 minutes and I had to suggest longer intervals. Another critique of the initial lesson plan was that the student gave a highly theoretical overview of complex statements without referring to the students’ real life experience. To improve upon the design and reduce lecturing the student constructed a true/false quiz (Figure 7) for the students to work through by way of introduction.

![Figure 7. The true/false quiz](image)

I found the use of police handcuffs a strange symbol, but the student explained the metaphor of two binary devices bound together and the rest of the group liked it, so the student went ahead.

![Figure 8. The main menu in English and Arabic](image)
The lesson was to be conducted in Arabic, but, because the student knew that I do not read or understand the language, she built a copy in English (Figure 8).

Once the learners had gone through the quiz and the concept had been established, they worked through a spreadsheet that allowed them to test various combinations, by way of self-assessment (Figure 9).

As in the evaluation course, students had to write a formal essay. It was assessed with a much less complex rubric than the essay described in the first scene, as it formed a much smaller component of the final grade. A much more intimate assessment was possible when we viewed the edited video produced by the student. Students used a small digital still and video camera with editing software. The camera was designed and marketed as a product to enable school children to produce their own video materials. It produced relatively low quality images, but created relatively small file sizes. The editing software that came with the camera was versatile yet intuitive. The use of this camera served two purposes, firstly it allowed for the assessment of the lessons, but possibly more importantly it was hoped that it would encourage the students to get their own learners to use the camera for projects later.

Images captured from the video (figure 10) show that the student was able to recognise and interpret various significant moments in the learning experience of the learners.

The student presented a 13 minute edited video complete with special visual and sound effects, transitions and subtitles. The subtitles included the name of the school, a description of the area being shown, such as the language laboratory and the computer laboratory, the activities taking place, such as initial installation of the software, briefing the students, and most importantly an interpretation of the events during the lesson.

The video showed that the student had been able to present a successful lesson in a computer-rich environment. In her talk accompanying the video the student explained what she had done during each phase of the lesson, and what happened. The student found it much less stressful to speak about the video playing behind her, than if she had been asked simply to talk about her lesson without any visual aids.
At a deeper level the video showed that the student was able to observe and pick out significant events during the presentation of the lesson. There was also evidence that she was able to tell a story visually, and had mastered the elements of video making. It was clear that she knew about establishing shots, mid shots and close-ups and when to use which. She had also mastered the elementary computer-based video editor and special effects generator and used those cleverly to emphasize certain events.

Over all, the case of the constructivist lesson shows a number of important aspects of portfolio assessment. Firstly, the assessment was initially formative. No grade was assigned and students were allowed to refine their work all the time to the final presentation. The aim of the assessment was to improve the quality of the students’ work, not to assign a grade. The assessment also attempted to be holistic in that the whole design cycle was assessed, and multiple instruments were used, ranging from authentic documentation in the form of the learning event protocol, through
analysis of artefacts produced for the lesson, to a formal essay assessment and an intimate self-evaluation of the video by the student.

Conclusions

This article set out to share ways in which a group of students and the author tried to reach commonality in assessing the quality of their learning across linguistic and cultural boundaries. The first scene showed that a positivist assessment had a high level of correlation between student and assessor grades, but tended to be mechanical and if students know exactly every point of the assessment, their products lack creativity. The second scene showed that the value of team productions were that students learnt from one another, while the downside still lay in a high level of “identical” work. Finally the third scene showed that a continuous, holistic assessment strategy, in which the entire process is monitored, allows us to obtain a rich, thick description of the learning journey, and a common understanding of what the student had learnt. Amin’s (2002) six factors formed the basis of developing commonality. By way of conclusion it would thus make sense to group the conclusions in terms of Amin’s six factors.

*Rapport with students* was crucial in obtaining their buy-in to the formative evaluation process. Negotiating the nature of the task and the criteria for assessment was a three-way process involving the instructor, the student and their peers. It was only possible to reach agreement when everybody was familiar with the situation of the person being assessed. Much of the rapport building took place outside class time (Cronje 2006) and falls outside the scope of this article, but a significant proportion of the rapport was created by way of *evaluation of assignments and feedback*.

Initially high anxiety levels among students meant that it was necessary to give regular feedback on small tasks, as is also suggested by MacKinnon & Manathunga (2003), who mentioned that students needed unambiguous instructions.

As time progressed it became necessary to withhold feedback to prevent students from becoming too dependent on the instructor and to encourage them to work independently.

Providing regular feedback on shorter tasks and negotiating tasks and assessment criteria with students on an almost individual basis helped regulate the workload of the students. The instructor came to know not only how competent students were, but how long it took them to become competent and how much work it took. Students were less likely to be swamped with work, as they had input into the process. However the workload of the instructor increased dramatically, as facilitating the course involved much more than preparing and presenting a traditional series of lectures and following it up with a test. However, the fact that lectures were almost excluded from the presentation mode freed up some instructor time. Nevertheless one should stress here that the scalability of this type of assessment is doubtful.

*Availability of Help* played an important part in ensuring the students’ success. Both the instructor and local facilitators provided help, but the process became unstuck on a number of times when institutional support was lacking during periods when I was not in Khartoum and we had to re-negotiate some tasks and criteria.

Working across radically different cultures means that *course organization* has to be done dynamically as both the instructor and the students become familiar with their altered situation. Much design flexibility is required. Once again a qualitative approach is useful here as the assignments provide evidence of the growth of the student, not just proof of meeting an external set of criteria.

I find it interesting that Amin speaks of *language or communications* rather than language and communications. In the course of this research it became very clear that communication was much more important than language. Often during discussions with students it became clear, from the animated nature of their discussion, that all was not well – although I could not understand a word of the Arabic, I could deduce from their body language, facial expressions and vocal tone that things were amiss. As time progressed it was even possible to determine the source of the problem, be it the nature of the task, university matters, interpersonal conflict, or whatever. Eventually the portfolio becomes evidence of a common understanding between all the participants involved in its creation.
Some of the concerns that emerge from the research relate to the mechanical nature of rubric-based assessment of portfolio items, which resonates with Johnston’s (2004) comment regarding the tendency of raters to self-moderate – thus detracting from the reliability of the results.

Another important aspect in qualitative assessment is the philosophical position of the assessor. It is my belief, and that of my two co-presenters, that these are good students. It was also the belief of their home institution who selected them. The purpose of our assessment was interpretive – to understand the students. The purpose of SUST’s assessment was positivist – to make the best possible and fairest (exclusive) selection. And we agreed. But, what would a completely external third party say, and would that differ if that third party were from the West or from the Arab-speaking region?

From the discussion and conclusions a number of recommendations arise.

**Recommendations**

Research is required to determine heuristics for the composition of credible portfolios, as suggested by Driessen, *et al.* (2005). As the heuristics develop, training should be developed for instructors and learners in the composition of portfolios. In the same way in which the skill of taking a test needs to be developed alongside the subject knowledge, the skill of composing a portfolio that reveals your skills and attitudes is a complementary skill that needs to be developed. In the same way that teachers prepare learners for an examination, they need to prepare learners for the assessment of their portfolios. In a cross-cultural context this becomes even more important as learners need to develop skills to let their true ability shine through the various cultural filters.

In this respect the role of off-task activities in building the rapport required for designing credible portfolios across cultures becomes important. Learners need to be evaluated in context, and that context extends beyond the walls of the classroom and beyond the traditional activities related to the subject matter.

Finally we need to investigate ways of making of qualitative assessment scalable. The cases presented here are drawn from a population of twelve. It would be impossible to give so much attention to a large group, yet, despite any group size each individual will be required to perform in a predictable way. I support Wilkerson & Lang’s (2003) caveat about using portfolios for purposes of exclusion, and agree with Driessen, *et al.* (2005) who believes the key lies in standardizing the procedures of establishing the audit trail and the dependability audit, rather than standardizing the procedures of the assessment itself, as the assessment should assess the learner, and the standardization should assess the assessment.

**Acknowledgement**

The contributions of Prof Seugnet Blignaut, Dr Dolf Steyn and the students who participated in this study are gratefully acknowledged.

**References**


