1-1-2014

Digital portfolios for summative assessment

Christopher P. Newhouse
Edith Cowan University, p.newhouse@ecu.edu.au

DIGITAL PORTFOLIOS FOR SUMMATIVE ASSESSMENT

C. Paul Newhouse
Associate Professor, School of Education, Edith Cowan University

p.newhouse@ecu.edu.au

Abstract

The collection and scoring of creative practical work for summative assessment across a large jurisdiction such as Western Australia is challenging. An alternative approach would be to submit digital representations as online portfolios. However, to give a valid and reliable measure the representations would need to be of adequate quality. Further, judgements of creative practical work are necessarily subjective giving concern about the reliability of scores for high-stakes assessment. The paired comparisons method of scoring lends itself to addressing this problem and is feasible where the work is in digital form. This paper reports on a three-year study to investigate the representation of student practical work in digital forms for the purpose of summative assessment in the Visual Arts and Design courses. This study set out to determine whether the digital approach was feasible and adequate fidelity could be achieved in order to use the paired comparisons method of scoring. The study found this process was feasible, and the results were acceptable. However, the approach lacked support from Visual Arts teachers and students who wanted the original artworks to be assessed. By contrast the attitudes and perceptions of Design teachers and students were very supportive.

Keywords: Visual Arts, Design, summative assessment, portfolio, digital representation

The summative assessment of performance on creative practical components of courses is difficult and therefore typically is either avoided or attempted in an inadequate manner. In particular the management and logistics of giving assessors access to the performance and the need for them to make highly subjective judgements reduces the reliability of the measurement (Dillon & Brown, 2006). It is likely that these challenges may be addressed through using digital technologies and comparative, rather than absolute, methods of scoring (Heldsinger & Humphry, 2010; Lin & Dwyer, 2006; McGaw, 2006; Pollitt, 2012). Further, if some of the obstacles to using portfolios for high-stakes assessment can be overcome by using digital technologies and modern psychometrics then this will better align assessment with preferred pedagogy (Clarke-Midura & Dede, 2010; Stobart & Eggen, 2012).

The assessment of creative practical performance has typically required students to submit a portfolio of work that includes created artefacts and process documents (e.g. Madeja, 2004). The assessor(s) judge this against a set of criteria to allocate a score or grade. This approach is problematic for widespread high-stakes implementation due to the difficulties in management and measurement (Clarke-Midura & Dede, 2010). This paper reports on some of the findings from the first two phases of a study that sought to address these difficulties through using digital technologies and the applications of modern psychometrics with portfolios in the *Visual Arts* and *Design* senior secondary school courses in Western Australia (W.A.).

Using portfolios for assessment purposes is not new but has had limited application for high-stakes summative assessment probably due to the costs of management and a lack of confidence in the reliability of the measurement. Masters (2013, p. 38) explains that, "When assembled over a period of time, portfolios can provide a valid basis for establishing current levels of achievement and for monitoring progress over time". Dillon and Brown (2006) point out that in the Arts the assessor needs to be provided with adequate evidence to make a balance of judgements between the technical and expressive quality of the work submitted. Thus the assessor needs to be given comprehensive access to the portfolio that is difficult when they are in physical forms but less so when they are in digital form. Dillon and Brown (2006) argue that digital technologies, "present the opportunity to capture, store,

and manage multiple forms of evidence about artistic products and processes" (p. 420). However, if a digital form is used then this must adequately represent the performance.

The scoring of portfolios for summative assessment purposes has typically employed an analytical method that requires assessors to make absolute judgements. Each assessable quality is described as a criterion with a set of quantified levels of performance, often represented in a rubric, and then use some form of Item Response Theory such as Rasch modeling to generate a score or grade (Humphry & Heldsinger, 2009). Although this is preferable to just adding up the scores on the criteria, psychometricians such as Pollitt (2004) argue that this will not accurately measure a student's "performance or ability" (p. 5) because the nature of the performance is holistic and therefore a holistic method such as the paired comparison judgement method is "intrinsically more valid". However, until recently this has not been feasible for large-scale assessment but may be with the use of purpose built software, computer networks and digital representations of performances (Pollitt, 2012). Applied to the assessment of creative work it could be argued that,

> *Advances in assessment theory, notably evidence-centered design (ECD) and new statistical techniques and technology tools for supporting the use of ECD in assessment development, are making the assessment of complex cognitive components that are exercised in multiple subject matter contexts much more feasible. ... Embedding assessments in digital learning systems opens up possibilities for assessing features that are important but that could not be measured reliably and efficiently in the past. (U.S. Department of Education, 2013, p. 53)*

The main purpose of the study was to determine the efficacy of digital representations of student artwork for the paired comparisons method of marking for the purposes of high-stakes summative assessment. The processes and resulting artefacts of student activity needed to be represented in digital form to measure their performance for summative assessment purposes. The study built on concerns that in courses where portfolio assessment is used judgements are not comparable between contexts, are not reliable due to the subjectivity of assessors, and are not cost-effective for large groups of students spread across large jurisdictions. The questions were whether adequate fidelity of representation could be achieved using common digital technologies in typical school settings and whether the paired comparisons method of scoring would provide adequate reliability for high-stakes summative assessment purposes.

Method

The study addressed questions concerning the adequacy of digital representation, scalability, and moderation and standard setting in three phases. In the first phase we digitised student work that was submitted for high-stakes summative assessment, in the second students digitised their own work to be submitted online for external assessment, and in the third we tested the feasibility of using our online scoring systems to support moderation and standard setting. The senior secondary W.A. courses of *Visual Arts* and *Design* were used as the contexts for the study. The study employed an action research design that involved the collection of a range of data analysed from the perspectives of students, teachers and assessors. Students were surveyed and interviewed, teachers and assessors were interviewed, and the scores from marking were analysed and compared. The analytical marking criteria provided for the course were used by the study and initially the technical specifications for digitizing the portfolios were determined through an analysis of syllabus requirements and a review of portfolios submitted in the previous year. For more details of these data collection processes refer to Newhouse and Tarricone (2014).

In Western Australia (W.A.) for many years the summative assessment in the senior secondary courses *Visual Arts* and *Design* has included analytically scoring a physical portfolio. For the *Design* course this was a collation of 15 A3 sheets of paper and for the *Visual Arts* course was a resolved artwork supported by a document containing an 'artist statement' and photographs of intended presentation of the artwork. All of these portfolios were required to be sent to a central location in Perth to be assessed

and then be returned to the students after they are scored. In a large jurisdiction such as W.A. this approach provides many logistical and management challenges to add to the limitations of measurement and the lack of an adequate enduring record for confirmatory purposes.

The first one-year *Development and Pilot* phase involved 75 *Visual Arts* portfolios from 10 schools and 82 *Design* portfolios from 6 schools being digitised by our research team and scored using both an analytical and a paired comparisons method. The efficacy of the digital representations was interrogated through interpreting the responses of students, teachers and assessors, and through a comparison with the scoring of the original physical portfolios. The sample was purposefully selected for experienced teachers who had taught the course for a few years; many were experienced assessors.

The second one-year *School-Based Implementation* phase involved 138 *Visual Arts* students from 13 schools and 110 *Design* students from 14 schools, with all students from the penultimate year of secondary schooling. This sample of schools was initially selected to ensure a representative range of typical schools were involved including two from country areas and some from each of the three school systems (government, Catholic and independent). Researchers supported teachers to facilitate students in digitizing their own portfolios and uploading these to an online repository. As Dillon and Brown (2006) suggest it was likely that with students digitizing their own work a more accurate representation would be formed.

A third *Moderation and standard setting* phase involving 15 *Visual Arts* teachers from W.A. country schools is underway. These teachers are being supported in the use of our online scoring tools for analytical marking and paired comparisons judging to view the digitised portfolios from the first phase of the study. The purpose is to determine the feasibility of teachers using the tools from anywhere with no face-to-face support to develop their understanding of the standards of work submitted and provide a form of moderation at a distance. At the time of writing this phase had not been completed.

Phase One – digitisation by researchers

From meetings with experts to review the course syllabus requirements, and examples of student work, a set of requirements for digitising were drawn up (refer to Table 1). Researchers were trained to digitise the portfolios at the central location to which all portfolios for W.A. had been delivered. The resulting files were checked, where necessary edited and uploaded to an online repository to be accessed by assessors. For details of the processes of digitisation and the results of implementation refer to Newhouse (2014). While all required files were created for each portfolio for *Visual Arts* it was not possible to fully implement the intended procedures and guidelines to ensure the highest fidelity of representation due to time and space constraints.

All digitised portfolios were scored using an analytical marking method and a paired comparisons judgement method facilitated by online systems. For the former we used a custom-built Filemaker Pro database system and for the latter a commercially available system called the Adaptive Comparative Judgements System (ACJS) (Pollitt, 2012). In addition the raw scores awarded for the original physical forms of the portfolios for the WACE (W.A. Certificate of Education) were obtained so that each portfolio had three scores (i.e. WACE, Analytical, and Pairs). For details of the processes of scoring, the assessment criteria and the analysis of the resulting scores refer to Newhouse (2014). The results of scoring indicated that assessor could readily use online tools for both methods of scoring for both courses and that the scores from paired comparisons judgement correlated strongly with the WACE scores. However, for both courses the inter-rater correlation for analytical marking was only moderate and for *Design* the scores from analytical marking did not correlate strongly with the other two scores. It was concluded that for *Visual Arts* the typically high level of subjectivity found in judging artworks could be counteracted by the combining of the judgements of multiple assessors, as best represented by the paired comparisons method. Whereas for *Design* this did not occur to the same extent because the scanned portfolios were too dense making it difficult for assessors to consistently locate information relevant to assessment criteria.

Table 1
Requirements for the digitising of the portfolios by the researchers.

Type	Digitisation Requirement	File type
Visual Arts 2 Dimensional	ID number and match-box visible in each photo/video	
	Photo of 'Artist Statement' and proposed installation if provided	.jpg
	One full size hi-resolution photo of 4 megapixels giving 300dpi at a reasonable size. Additional photos for multi-piece artworks.	.jpg
	4 x close ups - extracted from main photo(s)	.jpg
	All photos combined into one document	.pdf
	HD Video (pan & zoom) - 10 secs	.wmv
Visual Arts 3 Dimensional	ID number visible in each photo/video	
	Photo of 'Artist Statement' and proposed installation if provided	.jpg
	Full size photo + size object such as a match-box	.jpg
	At least 4 x angle photos (L, R, top, bottom)	.jpg
	4 x close ups - extracted from main photo	.jpg
	All photos combined into one document	.pdf
	HD Video (pan & zoom) - 10 seconds	.wmv
	3-D Animation for some works	.mov
Design	Colour scan of all A3 sheets	portfolio.pdf

Students and teachers viewed the digital representations of their own work, and then answered some questions. For details of the results of analysis of student and teacher responses refer to Newhouse (2014). From the survey of students it was clear that the two groups of students differed substantially in their previous experiences, skills, attitudes and perceptions regarding the digitisation of their work. In general the *Visual Arts* students were negative about using digital representations to replace marking of the physical artworks. They did not think the digital representations were adequate to mark in place of the original artworks. They had had very little previous experience in digitising their work and had less skill in using ICT than the *Design* students. In comparison the *Design* students were generally very positive about submitting work digitally, most had had some experience in doing so, and in fact most had produced some of their portfolio on computer. Unlike the Visual Arts students they would have liked to have a digital portfolio assessed, rather than a paper-based one, provided they could create and submit it digitally themselves. The two groups of teachers tended to mirror the attitudes and perceptions of their students in that the *Visual Arts* teachers were very negative about the potential of using digital representations for summative assessment whereas the *Design* teachers were very positive and perceived it to be more consistent with the intentions of the course and the realities of the relevant workplaces.

The first phase of the study demonstrated the potential for using common technologies for representing all the students' submissions as digital portfolios and scoring them using online tools. However, for both courses central digitisation was not feasible or desirable and it was likely that better portfolios could result from students creating and submitting their own equivalent digital portfolios. It appeared likely that *Design* students would have adequate ICT skills to accomplish this but it was likely that many *Visual Arts* students may not and therefore would need some support in doing so. This would be investigated in the second phase of the study. It was also determined that their were difficulties with the reliably scoring the *Design* portfolios probably due to the structure of the portfolios and the nature of the assessment criteria and therefore changes were recommended (Newhouse, 2014).

Phase Two – digitisation by students

For the second phase students were assisted to create a digital representation of their own work, similar to that for the first phase but with some enhancements. They used technical specifications based on those used in the first phase, however, for both courses they were given the opportunity to explain their work to the assessors through a video recording. For the *Design* portfolios (refer to Table 2) the students created one PDF file comprising 15 screens and for the *Visual Arts* portfolios students took their own photographs, including close-ups. This also meant that the *Design* portfolios could include considerably richer communication with more focus on images and layout than text, than was possible for paper submissions. Wherever possible students followed written instructions to create and submit their portfolios online. For more details of the processes involved in digitization refer to Newhouse and Tarricone (2014).

Table 2
Requirements for the Design digital portfolios produced by students.

Type	Digitisation Requirement	File type
Design	15 screens (guide is A3 size maximum)	
	Up to three projects included and only the best examples	portfolio.pdf
	PDF size should be approximately 12MB	
	Single PDF (Any software can be used to produce the portfolio as long as the maximum file size of 300dpi is not exceeded when saved as a PDF)	
	20 seconds of video supported by annotated audio descriptions focussed on the design process	video.avi
	Up to 12 MB in size	

All digital portfolios were scored using the two methods from the first phase, that is, analytical marking and paired comparisons judging.

For the Design portfolios, an analysis of the scores from analytical marking showed that although the two assessors had very similar means, standard deviations and ranges this only resulted in a moderate to low significant correlation between them (r=0.47, p<0.01). This would tend to indicate that the assessors were not very consistent between themselves and that the marking criteria were difficult to interpret consistently. Rasch analysis showed that whereas there was good fit to the model and good overall discrimination, with a person separation index of 0.94 and the Chi-square probability of 0.94, all item locations were on the lower half of the person locations, probably indicating that it was too easy to score highly on all criteria. For the Visual Arts portfolios there was a strong and significant correlation between the three assessors (r=0.73, 0.73, 0.75, p<0.01) and Rasch analysis found a high Cronbachs Alpha coefficient (0.96). However, this analysis did indicate the criteria were 'too easy' for the range of quality of the work.

The paired comparisons judging was facilitated by the *ACJS* online system. For the Design portfolios the Cronbach Alpha coefficient was 0.90 after the 13th round of marking that indicated a high level of reliability. In total it was estimated that assessors took from 4:33 to 9:22 minutes per judgement and that 689 judgements had taken on average 6:33 minutes per judgement. For the Visual Arts portfolios the reliability was also good (α=0.953). There was a strong correlation between the two sets of scores (analytical and pairs) for the *Visual Arts* portfolios (r=0.84, p<0.01) but only moderate for the *Design* portfolios (r=0.64, p<0.01). For Visual Arts this probably indicates that while the judgement of individual assessors is highly subjective their combined judgement is more reliable as represented by the paired comparisons scores and average of the analytical scores. It is likely that the lack of consistency in analytical marking of the *Design* portfolios was still mainly due to the quantity and complexity of information in the portfolios making it difficult to sample to make judgements.

Students and teachers were surveyed using questions similar to those used in the first phase. For more

details of the analysis of these data refer to Newhouse and Tarricone (2014). To a large extent their responses were similar to those in the first phase in that there was a contrast between the two courses with *Visual Arts* teachers and students generally negative and inexperienced and *Design* teachers and students generally positive and experienced.

The *Visual Arts* students generally were happy with the digital representations of their work. Of significance was that most indicated that the processes of digitisation and submission were easy to follow. They wrote that it was "easy", "fun" and "good to have a digital copy". Some had a number of pieces which made up one artwork and they found it helpful to capture all of the pieces together. They also considered that the close-ups showed the detail in their work and "emphasised tonal modulation". They felt that digitisation provided "quality clear images" and "good light, shadow and colour capture". Some commented that it would be useful for markers to know their 'reasons/points of view/hidden messages' about their work as provided in the annotated video. Despite this half would have preferred someone else to do the digitising, and most would still prefer an assessor to mark their original artwork, and in doing so they would receive a higher score. Some considered that digital images "hide faults" and "can make [the] artwork look better". Many indicated that the video was too short and that they did not enjoy talking on the video (Note: the video was of the artwork not of the student). Some felt that the digitised work "wasn't necessarily a true or good representation" of the original work and that their artwork would "have more impact and meaning if it was marked personally" by the markers as it "doesn't look impressive in digital form". They also felt concerned that digitisation did not show "fine details, texture and doesn't capture the essence" of the original artwork.

The *Design* students were not as positive as expected with some of this likely to be due to a perception that the video was not worth the effort and that collating the portfolio was time-consuming and in some cases difficult to locate the pieces. There were statements such as, "time consuming -- more confident with paper portfolio". However, in one school students wrote,

> *a lot easier and quicker to create*
> *show full understanding of design as the digital portfolio*
> *digital portfolio included everything*
> *happy with it*

Overall the *Visual Arts* teachers were still not in favour of digital representations being marked in place of the original artwork but were a little more positive. They perceived value in students being able to represent their work in this way and tended to be impressed with the photographs their students had taken. In addition two country teachers conceded that there were some logistical and security advantages to using digital portfolios in place of physical ones. One teacher explained the difficulties she had in transporting student work to Perth and commented that, "For the logistics of schools in this area, this is brilliant". Further, a Perth based teacher commented, "I have been very against this, but after some terminal damage to work for external examination I can see the benefit". Another teacher summed up the benefits,

> *There are TWO benefits with the digitalization process. One being it prevents student's work*
> *from the risk of being damaged during external assessment. The second benefit will be for*
> *teachers to have a record of all of their student's artworks. This is it's main potential and it is a*
> *process that can be done quite easily. This can also help with create portfolio of works that*
> *represent different grades, this could be used for training new markers and new teachers.*

Generally the *Design* teachers were in favour of students submitting digital portfolios although they recognized that many students would need to be better organised and improve some of their technical skills. A typical comment from a *Design* teacher was,

> *Most students were able to present their work really well after creating the first couple of pages.*
> *They looked much more professional than traditional cut & paste method. File management*

was the main problem for them but having gone through this process for the first time I would approach it differently for some students next time. Possibly creating the pdf at the end of the portfolio process. Merging all of their pages together.

Phase Three – moderation and standard setting

For the third phase the aim was to evaluate the suitability of our approach for implementation across the state for the purposes of moderation and standard setting by demonstrating that all teachers could be involved in scoring through our online systems. From the second phase we knew that students could upload their portfolios and that local Perth teachers could use the online systems supported by one or two face-to-face workshops. However, to be implemented statewide teachers from country areas would need to participate and thus all support would need to use online technologies. We could demonstrate the feasibility of this by recruiting a sample of country teachers to score the submissions from the first phase, identify the minimum support required to use our online systems to achieve results similar to those in the first phase. The phase is currently underway and not completed at the time of writing and therefore it is not possible to report results at this time.

Conclusions

At this stage our study has shown that visual artworks can be adequately digitised for the purposes of summative assessment and that students can do this using relatively inexpensive equipment, systems and software. However, teachers and most students in *Visual Arts* are not persuaded of the adequacy, although if faced without the choice of centralised marking of the physical forms of the work this may change. The lack of experience of students, probably because there is currently no need to represent their work digitally, is a likely explanation of the negative perceptions of many students. In comparison teachers and most students in *Design* are enthusiastic of the potential for digital submission of the portfolios. The limitation to static text and graphics meant that there was little difference between the digital and paper representations and the focus and structure of the existing portfolio appeared to make it difficult to mark because there was too much information to synthesise across too many projects and in too many variations of layout and location. Therefore the content of the portfolio would need to be reviewed to gain the affordances of being a digital portfolio.

In Australia as the trend continues towards a national curriculum with accountability requiring comparability of assessment it is almost certain that online judgement or marking systems will need to be used to be cost-effective. This approach will improve manageability (e.g. marking from anywhere, less use of physical space and time), increase the reliability of the scores from marking, maintain an enduring record, and provide knowledge of assessor perceptions. Using various forms of digital portfolios this can probably be achieved in any curriculum area, even difficult areas such as visual arts. Our study is demonstrating the viability of the technology so now political will and informed community attitudes are needed to make the key decisions to move all aspects of assessment into the digital age.

Acknowledgement

References

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education, 42*(3), 309-328.

Dillon, S. C., & Brown, A. R. (2006). The art of ePortfolios: insights from the creative arts experience. In A. Jafari & C. Kaufman (Eds.), *Handbook of Research on ePortfolios* (pp. 420-433). Hershey PA: Idea Group Inc.

Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher, 37*(2), 1-20.

Humphry, S., & Heldsinger, S. (2009). *Do rubrics help to inform and direct teaching practice?* Paper presented at the Assessment and Student Learning: Collecting, interpreting and using data to inform teaching., Perth, Western Australia.

Lin, H., & Dwyer, F. (2006). The fingertip effects of computer-based assessment in education. *TechTrends, 50*(6), 27-31.

Madeja, S. S. (2004). Alternative assessment strategies for schools. *Arts Education Policy Review, 105*(5), 3-13.

Masters, G. N. (2013). Reforming educational assessment: Imperatives, principles and challenges. In S. Mellor (Series Ed.). *Australian Education Review*. Melbourne, Victoria: ACER Press.

McGaw, B. (2006). Assessment to fit for purpose. In conference proceedings, *32nd Annual Conference of the International Association for Educational Assessment* (pp. 1-16). Singapore: International Association for Educational Assessment.

Newhouse, C. P. (2014). Using digital representations of practical production work for summative assessment. *Assessment in Education: Principles, Policy & Practice, 21*(2), 205-220. doi: 10.1080/0969594X.2013.868341

Newhouse, C. P., & Tarricone, P. (2014). Digitising practical production work for high-stakes assessments. *Canadian Journal of Learning and Technology, in press*.

Pollitt, A. (2004). *Let's stop marking exams.* Paper presented at the International Association for Educational Assessment Conference, Philadelphia, June 2004. Paper retrieved June 24, 2007 from www.cambridgeassessment.org.uk/ca/Our_Services/Research/Conference_Papers

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*(3), 281-300.

Stobart, G., & Eggen, T. (2012). High-stakes testing - value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice, 19*(1), 1-6.

U.S. Department of Education. (2013). *Expanding Evidence Approaches for Learning in a Digital World.* Washington, D.C.: U.S. Department of Education, Office of Educational Technology.