Research and Practice in Technology Enhanced Learning
Vol. 9, No. 3 (2014) 475－492

USING DIGITAL PORTFOLIOS FOR HIGH-STAKES ASSESSMENT IN VISUAL ARTS

C. PAUL NEWHOUSE

School of Education, Edith Cowan University, Australia
p.newhouse@ecu.edu.au

The collection and scoring of student artwork, for high-stakes assessment, across a large jurisdiction such as Western Australia is challenging. An alternative approach would be to submit digital representations of the artworks online for assessing. However, to give a valid and reliable measure the representations would need to be of adequate quality. Further, judgements of artworks are necessarily subjective giving concern about the reliability of scoring. The comparative pairs method of scoring lends itself to addressing this problem and is feasible where the work is in digital form. This paper reports on one component of a three-year study to investigate the representation of student practical work in digital forms for the purpose of summative assessment. The first phase of the project involved the researchers creating digital representations of the artwork submitted at the end of secondary schooling by a sample of students. The second phase involved a sample of students creating digital representations of their own work and submitting them through an online system. The study found this process was feasible, and the results were acceptable, but it lacked support from teachers and students who wanted the original artworks to be assessed. The comparative pairs method of scoring was found to provide reliable scores and be well suited to making judgements around creative work.

Keywords: Visual Arts; summative assessment; portfolio; digital representation.

1. Introduction

Creative expression is difficult to assess largely due to the subjective nature of the judgements made (Dillon & Brown, 2006). Where these judgements contribute to high-stakes assessment there are concerns about the reliability of resulting scores. Recent research has suggested that a comparative pairs method of marking may be the best approach to counter the subjectivity of judgements (Heldsinger & Humphry, 2010; Kimbell, Wheeler, Miller, & Pollitt, 2007; Newhouse, 2011; Pollitt, 2012). This method is based on making comparisons between pairs of artefacts or representations of performance and combining the results of these comparisons using a Rasch dichotomous model (Pollitt, 2012). However, to effectively apply this method to a large sample realistically requires that the work being judged is in a digital form so that computer software and networks can be used to access the work, enter judgements and calculate scores. For the results to be valid these digital representations of student work or performance must have an adequate level of fidelity for the purposes of the assessment. At the same time the creation and submission of these digital representations must be

manageable and at minimal cost. Can this be accomplished for senior secondary school high-stakes summative assessment for a highly creative area of the curriculum?

This paper reports on a component of a three-year study that sought to address this question. This study used the practical work in two senior secondary courses in Western Australia (W.A.), Visual Arts and Design, however, this paper only considers findings associated with the Visual Arts course. The study built on five-years of research into the use of digital technologies to support high-stakes summative assessment, which had included investigation of the comparative pairs method of marking in collaboration with researchers from the British e-scape project (Kimbell et al., 2007; Newhouse & Njiru, 2009). The main purpose was to determine the efficacy of digital representations of student artwork for the comparative pairs method of marking for the purposes of high-stakes summative assessment. The study built on three main intersecting areas of knowledge: assessment; psychometrics; and digital representation. The processes and resulting artefacts (e.g. paintings and sculptures) of student activity needed to be represented in digital form to measure their performance for summative assessment purposes.

The assessment of practical creative expression, such as for Visual Arts, has typically been done by students submitting a portfolio of work that includes created artefacts and process documents (e.g. Madeja, 2004). This portfolio is then judged by an expert against a set of criteria and awarded a score or grade. This approach has been reasonably effective for small-scale formative assessment purposes such as with a teacher and her class. However, when the scale is larger and the stakes are higher, management and measurement reliability become obstacles (Clarke-Midura & Dede, 2010). This study sought to address management obstacles through using digital technologies and reliability obstacles through applying modern psychometrics (Clarke-Midura & Dede, 2010; McGaw, 2006; Ridgway, McCusker, & Pead, 2004).

Using portfolios for assessment is part of what Messick (1994) refers to as "performance-and-product assessment" (p. 14) where a performance concerns processes and a product is a remaining outcome. In the Visual Arts course in W.A. the existing focus of the final assessment was on the product, with the process represented in a minor subsidiary form, although some would argue it should be as important as the product (Dillon & Brown, 2006). Dillon and Brown point out that in the visual arts while the product may be "tangible" the "meaning may not be clear or literal" (p. 421) so this needs to be captured in representations of the process. They also highlight the assessment problem of creating "representational and evaluative" frameworks that help identify differences in both "technical and expressive ability". That is so the assessor is provided with adequate evidence to make a balance of judgements between the technical and expressive quality of the work submitted, as representative of the student's "artistic" knowledge. That is why in our project we used a combination of photographs, video, audio and text to represent the student's work, with the artist's "voice" (p. 419) in the audio and text providing some explanation of the process and meaning.

Traditionally portfolios of creative work have been assessed for summative purposes using a range of analytical techniques that aim to quantify identifiable aspects or qualities of the work. This is much easier to accomplish for technical qualities than expressive (creative) qualities that are difficult to describe alphanumerically and detached from the assessor (Dillon & Brown, 2006). Psychometrics is the science that focuses on measuring such mental processes, quantifying the qualitative (Barrett, 2003). An analytical technique used by psychometricians is to describe each assessable quality as a criterion with a set of quantified levels of performance or achievement, 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 comparative pairs 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 (McGaw, 2006; Pollitt, 2012).

Dillon and Brown (2006) argue that, "Digital media and information systems present the opportunity to capture, store, and manage multiple forms of evidence about artistic products and processes" (p. 420). The question is how can the knowledge and skills being assessed best be represented in digital media to allow comparative judgements to be made? Then the question is whether this can be accomplished in typical schools with their accompanying constraints? Our study aimed to address these two questions as part of an overarching research question: In what ways may the digital representation of Visual Arts portfolios and the use of the comparative pairs method of marking lead to better comparability of scores between work done in different media and contexts and better alignment with student capability?

2. Method

Our study addressed the questions of adequate digital representation and scalability to schools in two phases. In the first phase we digitised student work that was submitted for high-stakes summative assessment, and in the second students digitised their own work to be submitted online for external assessment. In the first phase we compared the results of marking the digitised artwork portfolio with the results of marking the physical artwork portfolio. In Western Australia (W.A.) for many years Visual Arts in senior secondary schooling has been assessed through analytically judging a physical portfolio that at one stage was more a representation of process, termed a "visual diary", but more recently is largely the physical artworks themselves supported by a document containing an "artist statement" and photographs of intended presentation of the artwork. 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 study employed an action-research evaluation design with the two development-evaluation phases 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" (either in person or by email), and the scores from assessing were analysed and compared. The analytical marking criteria provided for the course were used by the study and initially the technical specifications for digitising the portfolios were determined through an analysis of syllabus requirements and a review of portfolios submitted in the previous year.

The first one-year *Development and Pilot* phase occurred from July 2011 to June 2012, and involved 75 Visual Arts portfolios from 11 schools being digitised by our research team and marked using both an analytical and a comparative pairs method. The efficacy of the digital representations was interrogated through interpreting the responses of students, teachers and assessors, and through a comparison with the scores awarded to the original physical portfolios (Note: The awarding body provided us with the raw scores from the official marking for the students in our sample.). The sample was purposefully selected to ensure all teachers were experienced in teaching the *Visual Arts* course, having taught the course for a few years; many were also experienced external assessors.

The second one-year *School-Based Implementation* phase occurred between July 2012 and June 2013, and involved 138 students from the penultimate year of 13 secondary schools. 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 digitising their own portfolios and uploading these to an online repository. As Dillon and Brown (2006) suggested it was likely that with students digitising their own work a more accurate representation would be formed. This is partly because the process of creating the digital artefacts added to the student's "voice" or communication through the artwork, that is, the act of digitization is part of the creation of the artwork. For this reason they assert, "Students must maintain their creative integrity and input into the selection of the ePortfolio content …" (p. 427).

3. Phase One – Digitisation by Researchers

In this phase of the study the artworks submitted by final year secondary students in the sample for final examination were represented digitally by the researchers to allow online scoring. This was a separate set of processes from the official scoring of the artworks for the purposes of graduation and tertiary education entrance. Prior to digitisation a set of procedures and guidelines for digitising were drawn up after consultation with experts (awarding body curriculum officers and University academics from the area) to review the course syllabus requirements and 106 examples of student two or three-dimensional (2D or 3D) artworks from an exhibition. These procedures and guidelines were then tested with work produced by students from a Year 11 class at a local school and refined to give those listed in Table 1. Three teams of researchers were trained in using SLR

Table 1. Procedures and guidelines for the digitising of the Visual Arts portfolios by the researchers.

Artwork type	Digitisation Requirement	File type
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 seconds	.wmv
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

(Single Lens Reflex – these generally allow better quality photographs) and digital video cameras to digitise the portfolios at the central location to which all artwork portfolios for the state of W.A. had been delivered. Unfortunately, it was not possible to fully implement the procedures and guidelines as intended because we were only given one day to access the work and the location had very little space available. Therefore backdrops and lighting could not be set up, most 3D artworks could not be moved, and time did not permit proper colour balancing and multiple attempts at photographs and videos. However, despite the constraints a full set of digital files was created for each artwork portfolio. Each portfolio included either 2D or 3D artwork, almost all included a one-page artist statement and most included a photograph of the intended presentation. A substantial number of the artworks comprised multiple pieces with the maximum number being 11 interconnected two-dimensional paintings.

Prior to uploading the digital files into the online repositories to be accessed by the assessors some simple maintenance was required. This included rotating and cropping photographs, and for the videos a change of file format to WMV. The four close-up images were created from the original photographs based on advice from a Visual Arts teaching specialist. In addition a single PDF file was created using all the original photographs and the close-up images. All the files for each student were copied onto USB flash drives and given to them to review prior to completing a questionnaire. This process also checked that the correct files were associated with each student prior to all files being copied to a server for analytical marking, and uploaded to the MAPS online portfolio system for comparative pairs judging.

Three experienced assessors (had been official markers for the physical artworks in the previous year) were employed for analytical marking of the digital representations and they were augmented with the teachers and some officers of the awarding body for

comparative pairs judging. The scoring criteria for analytical marking of the digital representations were those used for the official marking of the physical artwork portfolios. The criteria are listed below along with descriptors for the highest level of achievement for each (possible range of score-points are in parentheses).

C1: Creativity and innovation: Artwork/s is outstanding, showing exceptional creativity and innovation and the emergence of a distinctive style. (0-6)

C2: Communication of ideas: Ideas are skilfully realised and powerfully communicated in sophisticated and highly coherent resolved artwork/s. (0-5)

C3: Use of visual language: Extensive and sophisticated application of visual language in the artwork/s. Complex and highly resolved visual relationships are evident. (0-12)

C4: Use of media and/or materials: Highly discerning selection and refined use of media and/or materials demonstrating sensitive application and handling. (0-5)

C5: Use of skills and/or processes: Extensive and sophisticated selection and application of skills and processes. (0-12)

From these criteria one holistic criterion, to be used in the comparative pairs judging of the digital representations, was generated at a workshop involving all the assessors. Because comparative pairs judging is by nature holistic, in their minds assessors must distil the criteria they use into a single one, and thus it is useful to do this explicitly with all assessors to promote consistency of judgement. This holistic criterion was stated as,

Judgement about performance addresses students' ability to creatively use visual language, materials and processes to skilfully communicate an innovative idea in a resolved artwork.

Both methods of scoring the digital representations were facilitated through online systems accessed through a standard Internet browser. For analytical marking the research team developed a database system using *Filemaker Pro* that displayed the digital representations and the rubric, and allowed scores to be recorded using on-screen buttons. For comparative pairs judging the *Adaptive Comparative Judgements System* (ACJS) associated with the MAPS online portfolio system was used. This system, as described by Pollitt (2012), facilitates all aspects of the comparative pairs method. This method involves assessors being asked to select "winners", based on a holistic criterion, between successive pairs of portfolios (for a more detailed explanation refer to Pollitt (2012)). All assessors were trained in the use of the system at the workshop, where in addition, the holistic criterion was generated.

Teachers were asked to view the digital representations of their own students' work, provide a ranking, and then answer some questions. Overall they were negative about using digital representations to replace marking of the physical artworks. They believed that assessors needed to be able to "touch" the artwork and get a feel for the "size" of the work. Further they believed that some artwork would not photograph well due to the materials and techniques used and that three-dimensional work would not be adequately represented. From the survey of students (n=70) it was also clear that they did not think the digital representations were adequate to mark in place of the original artworks.

However, an analysis of the resulting scores showed a high significant correlation between scores from both methods of marking the digital representations and the official scores awarded for the original physical forms ($r=0.86$ for analytical and $r=0.74$ for comparative pairs, $p<0.01$). In fact these correlation coefficients were much larger than the inter-rater correlation coefficients between the three analytical assessors of the digital representations ($r=0.51$, 0.54 and 0.56, $p<0.01$). Almost identical correlation coefficients were found between rankings of portfolios as between the scores.

4. Phase Two – Digitisation by Students

For the second phase students were assisted to create a digital representation of their own work, in a similar manner to that employed by the researchers for the first phase of the study. They used the technical specifications shown in Table 2, with the only major changes from the first phase being, the opportunity to explain their work to the assessors as they recorded their video recording, and their choice of the close-up images photographed directly (not created digitally after capture). They were instructed to have a 2D or 3D artwork ready and they were guided on how to use the SLR digital and video cameras prior to digitisation. To effectively digitise the artworks a backdrop support and lighting kit was used with the cameras. The ePhoto kit (ePhoto Inc., 2012) used was inexpensive at less than $150 per kit and was purchased online (Figure 1). It included two light stands and umbrellas, two 45w day light bulbs (5500K/2 x 32”), one black and one

Table 2. Technical specifications provided for Visual Arts students and teachers (abridged version).

Component of digital representation	File type
ARTIST STATEMENT	
300 words containing the rationale for the conceptual and material development	.doc
DIGITAL PHOTOGRAPHS	
Backdrop stand and sheet is to be used for camera/video lighting.	
Room natural or artificial light, ensuring that there are no shadows.	
Photos need to match the colour of the artwork.	
Ruler needs to be visible to indicate size of the work.	
Photographs need to be in focus.	
High resolution, 4 megapixels to give 300dpi. 1,000 pixels on longest side and/or up to 5MB in size.	
MAIN PHOTO – 2D	
2D artwork needs to be completely upright.	.jpg
1 full size main photo of the work using a ruler as a size guide (photo 1).	
If the photo does not capture the work, a second photo can be taken.	
MAIN PHOTO – 3D	
1 full size photo taken from the front of the work including a ruler as a size guide	
4 full sized angle photos (left, right, top, bottom)	
CLOSE UP PHOTOS	
2D – 2 photos of own choice to highlight two aspects of the work	.jpg
3D – 4 photos of own choice to highlight four aspects of the work	
VIDEO – 2D & 3 D	
Up to 12MB of HD Video (pan and zoom) - 20 seconds with audio annotations	.avi

Figure 1. Backdrop, stand and lighting kit system (ePhoto Inc., 2012).

white muslin backdrop, and one background support system. As the system was used in Australia, the four light bulbs that came with the kit were replaced with 105W 5500K bulbs and light holders. The backdrop and lighting system was easy to store, transport, set up and use. Students could easily be instructed on how to set up the system and their work for digitisation, and uploading the files to the MAPS online repository.

Students completed a questionnaire (n=131 out of 142 students involved) about the creation of the representations of their artwork, and their attitudes towards, and perceptions of, the digital representation of their artwork. The questionnaire consisted of 25 closed response items and three open-response items. Standard descriptive statistics were generated for the closed response items. It was clear from their responses that they had little previous experience in representing their artwork digitally with 86% indicating little or no experience. Further, 79% indicated that previously none of their work had been submitted in digital form for any form of assessment. In fact, very little of their work had been originally created on computer but 55% indicated that they needed little or no help to digitise their artwork. They were split evenly on confidence in digitizing, but 70% thought it was easy to do the digitisation and 81% believed the requirements were easy to follow. The majority (62%) felt that the photographs and video represented their artwork well. However, 55% would have preferred someone else to do the digitising, 87% would still prefer an assessor to mark their original artwork, and 80% believed that in doing so they would receive a higher score. In some ways these response patterns

appear contradictory, with the digital representations perceived to have adequate fidelity but a perception that assessors would not recognize this, and a relatively high level of confidence in digitization, given a lack of experience, but a preference to forfeit control. Student confidence and perceptions were considered important factors in determining how digital representations could be used for summative assessment.

The responses to sets of closed-response items were combined to create three scales. Descriptive statistics for these scales are presented in Table 3 and the distribution of scores in Figure 2. Scores on the *eAssess* scale that represented their attitude and perception of the efficacy of the digitisation, indicated quite a positive view with the mean above the mid-point of 2.5. However, with a minimum of 1.6 and a substantial proportion below 2.5 there were clearly many who harboured some dissatisfaction. At school they used computers for an average of 60 minutes per day (mean on *SCUse* scale). Overall they indicated a fairly good self-assessment of their computer skills with a mean on the *Skills* scale of 2.9, well above the mid-point of 2.5. About half of their scores were spread between moderately (score of 3) and highly (score of 4) competent (Figure 2). Of the types computer applications listed in the questionnaire, students felt most confident about digital photography and image editing software (80% and 77%, consecutively, with excellent skills) while they felt least confident in using video editing software (20% with no skills).

There were open-response items that allowed them to identify the best and worst things about creating the digital representations of their work. Most stated that the

Table 3. Descriptive statistics for the scales based on items from the Visual Arts student questionnaire.

	N	Min	Max	Mean	SD	α	Description
eAssess	126	1.6	3.5	2.6	0.3	0.65	Efficacy of the digitisation. Score between 1 and 4
Skills	126	1.6	4.0	2.9	0.6	0.76	Self-assessment of ICT skills. Score between 1 and 4.
SCUse	131	0	480.0	60.1	68.5	na	Estimate of time in mins/day using computers at school.

α = Cronbach Alpha reliability coefficient

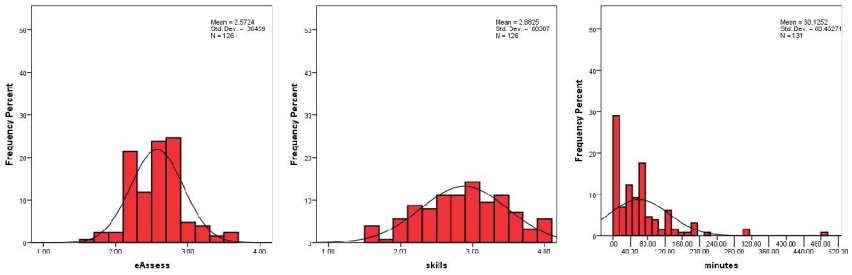

Figure 2. Graphs for the distribution of scores for the scales on the Visual Arts student questionnaire.

process was "easy" and "fun" and they thought it was good to have digital copies of their work. Some had a number of pieces that made up one artwork and they found it helpful to capture all of the pieces together. They also liked being able to take close-up photographs and talk on their video to help them foreground details of their work. Most were happy with the quality of the photographs even to the extent that some thought they could make the work look better than it really was. Many comments also indicated that they thought the digital versions would be easier to mark and would save on transportation, storage and damage. However, there were a number of things about the digital representations that they did not like, such as, that the video was too short and that they did not enjoy talking on the video. Some argued that their original work would have "more impact" and look more "impressive" and that digitisation wasn't a true representation. Some also felt concerned that digitisation did not show "fine details, texture and doesn't capture the essence" of the original artwork.

Overall the teachers were still not in favour of digital representations being marked in place of the original artwork but some made positive comments about the outcomes of the digitisation by their own students. One teacher stated: "I was very against it [digitisation] but after some terminal damage to work for external examination I can see the benefit from a wear and tear point of view". They were more impressed with the quality of the photographs but still felt they missed some "subtle nuances". Generally they believed that the original work would score higher because the viewer could "interact" with the work, the scale of the work was more obvious, and that the "digital representations do tend to flatten artwork and work seems to lose its tactile quality".

5. Scoring the Digitized Artwork Submissions

The digital "portfolios" from the second phase were scored using the same two methods used in the first phase: analytical "rubric" marking; and comparative pairs judging. However, because these portfolios were submitted from Year 11 students they were not for an official score as was the case for the first phase (Newhouse, 2014).

For both methods of scoring of the portfolios assessors used online digital tools through standard Internet browsers that accessed files from digital repositories. Three assessors scored the portfolios using a custom-built online database system that facilitated an analytical marking method based on a rubric. The scores for the criteria were totalled and then averaged between the three assessors. They were already familiar with the criteria so after a little practice with the tool they were able to access the system from workplaces and home to enter their marks.

Later, all work was scored again using a comparative pairs method that was facilitated by the ACJS online system associated with the MAPS portfolio system. Digital representation files had to be uploaded by a research assistant as separate journal entries in the *MAPS* system, to be transferred to the ACJS by staff of TAG Learning (the owner of MAPS). After a little practice with the system seven assessors (including the three involved in analytical marking) were able to access the system from workplaces and home to enter their judgements. This was sometimes a slow process where large files had

to be downloaded. Making the judgements was generally quick and convenient with many of the comparisons able to be made without looking at all files (i.e. where the two sets of work were very different in quality. The reliability of the results of comparative pairs marking was good (α=0.953).

An initial half day workshop was conducted to introduce the comparative pairs method of marking, develop and agree on the holistic criterion for this method of marking, and learn to use the AJCS. The holistic criterion was developed based on those from the WACE practical submission used in the analytical marking. There was some discussion of the need to make a holistic judgement but keep in tension criteria related to technical capability and creative expression. It was recognised that this would be difficult to achieve. Assessors completed some judgements during the workshop and the remaining judgements were completed over the next four weeks with assessors working from home or workplaces.

In general, for both methods of marking the assessors had little trouble accessing and using the online tools whether at home, school or University. The ease of storage, backup, transmission, access and sharing of portfolios in digital form were readily seen as advantages for the processes of assessment. With regard to the analytical marking, assessors found it convenient to view the work sample and the marking rubric, and to enter their marks on the same screen. Further, the database tool recorded and summed the scores that reduced the potential for error and made the assessors work easier. Assessors found the comparative pairs AJCS tool to be "very easy to use" and a "user friendly process". Even so, many found that there were "delays" and that it was "slow" to download files. This was mainly due to file size. One assessor called it "clunky" and the tool "seriously impaired the whole process of objective analysis", and it did not work in all browsers. Some even considered the system to be "primitive" (it should be noted that we were using a development version of the system). The assessors would have found it "helpful to view both images clearly, side by side on the computer screen, to make the final comparisons". There was also zoom in/out had limited functionality making it difficult to view the whole artwork on the screen. Assessors found it "very easy" to enter judgements.

The amount of time taken by the three expert assessors in the analytical marking varied from about 4 minutes to a maximum of 20 minutes per student. They took a total of 23 hours for the 142 submissions that is an average of 9.7 minutes per submission per assessor. The longer times were associated with student work that had more components and required larger files to be downloaded. The comparative pairs marking involved 7 judges. The ACJS estimated that 914 judgements had taken 54 hours and 15 minutes at 3:33 minutes per judgement. This was likely to be an overestimate because time includes if a judge takes a break while on a judgement. Judges ranged from 1:15 to 9:21 minutes per judgement (the former making 99 judgements and the latter only 20).

6. Results of Scoring the Digitized Artwork Submissions

To discuss the results from analytical marking we will simplify nomenclature by using abbreviations to represent the five marking criteria for the Visual Arts portfolio (the range of possible score-points is indicated in parentheses).

C1: Creativity and Innovation (0-5)
C2: Communication of Ideas (0-5)
C3: Use of Visual Language (0-10)
C4: Use of Media and/or Materials (0-8)
C5: Use of Skills and/or Processes (0-10)

Summaries of some of the descriptive statistics on the results of this marking are provided in Table 4. Despite the differences between the minimum scores given by the three assessors, there was no significant difference between their average scores. A maximum possible score of 38 was given by all three assessors. Assessor 2 appeared to have given the least varied scores with a minimum of 11 and a standard deviation of 6.3 while Assessor 3 gave the most varied scores with a minimum of 3 and a standard deviation of 7.9. Consistently, across all criteria, the highest scores were awarded by Assessor 2, and there were significant differences between the mean scores given by the assessors on the criteria. However, across the five criteria the average scores between the assessors were remarkable consistent (56%-60%).

Further Rasch analysis found a high Cronbach's Alpha coefficient (0.96) indicating that the scores resulting from this method were a reliable measure for this assessment. However, this analysis did indicate the criteria were "too easy" for the range of quality of the work. Correlation coefficients between the scores generated by the three assessors are shown in Table 5. This analysis of the scores shows a moderate to strong and significant

Table 4. Descriptive statistics from the analytical marking of the portfolios in the second phase of the study.

		Mean (SD)					
(N=142)	Range	Total	C1	C2	C3	C4	C5
Assessor1	5.0-35.0	20.7 (6.4)	2.9 (0.9)	2.6 (0.9)	5.6 (1.8)	4.1 (1.5)	5.5 (1.8)
Assessor2	11.0-38.0	24.8 (6.3)	3.3 (0.8)	3.4 (0.9)	6.4 (1.8)	5.4 (1.4)	6.4 (1.8)
Assessor3	3.0-38.0	19.7 (7.9)	2.9 (1.1)	2.9 (1.0)	5.1 (2.0)	4.0 (1.8)	4.8 (2.3)
Average	8.0-35.0	21.7 (6.2)	3.0 (0.8)	3.0 (0.8)	5.7 (1.7)	4.5 (1.4)	5.6 (1.8)
Average (%)	20.0-87.5	54.3 (15.5)	60.0 (16)	60.0 (16)	57.0 (17)	56.2 (18)	56.0 (18)

Table 5. Correlation coefficients between scores from assessors using analytical marking of the portfolios for the second phase of the study.

(N=142)	Assessor1	Assessor2	Assessor3	Average
Assessor1	1.00	0.73**	0.75**	0.91**
Assessor2		1.00	0.73**	0.89**
Assessor3			1.00	0.92**
Average				1.00

**. Correlation is significant at the 0.01 level (2-tailed).

correlation between the three assessors with correlation coefficients all approximately 0.73 ($p<0.01$) for both scores and ranking. This indicates that the scores were consistent between assessors and that the assessors used the marking criteria consistently. This was a better outcome than in the first phase of the study where the correlation coefficients were all around 0.54.

There were 13 schools involved in the second phase of the study. The data were analysed separately for each school as a case study. A summary of means and standard deviations from external analytical marking for the case studies is provided in Table 6. The rank for the average of the external assessors marks was for all students in the sample across all classes. For example, an average assessor rank of 46.5 implies that this student's total score ranked equal 46th out of all students. Not surprisingly there were substantial differences in means between the schools with the top mean for the five students from VJ and the lowest for the six students from VU. Once again it was remarkable how consistent the average scores were across the five criteria and 12 schools. That is, a high scoring school such as VJ, on average scored similarly highly on all five criteria, and a low scoring school such as VU, did so for all five criteria. This would tend to indicate that overall the assessors were tending to make a holistic judgement of an artwork and then using this to guide the allocation of scores for each criteria.

It had been decided to stop judging using the ACJS once the reliability coefficient (equivalent to Cronbach's Alpha) was at least 0.95. This occurred after the 13th round of judging when this coefficient was 0.953 and indicated a reliable set of scores had been produced. The *ACJS* provided a summary of results that included for each portfolio a "Parameter" score, a Standard Error (SE), an Unweighted mean square, an Unweighted

Table 6. Tabulation of results from marking as means (standard deviations) for each case in the second phase of the study for the Visual Arts course for the second phase of the study.

		Mean (SD) for criteria scores					Mean (SD) for total score	
Case	N	C1 (5)	C2 (5)	C3 (10)	C4 (8)	C5 (10)	Score (%)	Rank
VA	5	2.6 (0.9)	2.7 (1.0)	5.1 (1.6)	3.9 (1.3)	5.1 (1.8)	51.1 (17.1)	83.4 (44.2)
VB	6	2.6 (0.5)	2.6 (0.5)	4.9 (0.7)	4.3 (0.7)	5.2 (0.8)	51.6 (7.2)	86.4 (22.9)
VG	17	3.6 (0.6)	3.7 (0.5)	7.1 (1.1)	5.9 (1.0)	7.3 (1.3)	72.6 (11.3)	31.3 (24.9)
VH	13	2.4 (0.6)	2.3 (0.7)	3.9 (1.4)	2.9 (0.8)	3.6 (1.2)	39.9 (11.4)	113.3 (23.1)
VJ	5	3.9 (0.3)	4.0 (0.7)	7.7 (0.9)	5.9 (0.9)	7.7 (0.7)	76.8 (8.2)	22.4 (16.4)
VL	9	3.4 (0.6)	3.4 (0.8)	6.9 (1.1)	5.5 (1.1)	6.6 (1.4)	67.7 (12.3)	44.7 (31.4)
VM	11	3.1 (0.6)	3.1 (0.8)	6.4 (1.3)	5.0 (1.1)	6.4 (1.3)	63.1 (12.9)	56.5 (35.6)
VO	4	2.9 (0.3)	2.6 (0.2)	4.6 (0.4)	3.2 (0.3)	4.6 (0.3)	47.0 (3.7)	101.9 (11.2)
VP	37	2.7 (0.6)	2.7 (0.6)	5.1 (1.2)	3.7 (1.0)	4.7 (1.3)	50.0 (11.8)	90.4 (29.8)
VQ	8	3.0 (1.4)	3.2 (1.0)	6.4 (1.8)	5.1 (1.2)	6.4 (1.5)	63.6 (16.9)	58.7 (43.2)
VS	14	3.2 (0.8)	3.1 (0.8)	5.9 (1.9)	4.7 (1.5)	5.8 (2.0)	59.5 (17.9)	62.0 (42.9)
VU	6	2.4 (0.5)	2.3 (0.4)	4.1 (1.1)	3.2 (0.7)	3.9 (0.8)	41.7 (8.6)	114.3 (19.8)
VY	7	3.6 (0.5)	3.5 (0.6)	6.6 (1.1)	5.6 (1.5)	6.8 (1.7)	68.7 (14.2)	42.9 (39.3)
All	**142**	**3.0 (0.8)**	**3.0 (0.8)**	**5.7 (1.7)**	**4.5 (1.4)**	**5.6 (1.8)**	**57.2 (16.4)**	**71.0**

*Ranking of external assessors is for all 142 students involved in the study.

Z-score, a Weighted mean square, and a Weighted Z-score. Nine portfolios had an SE above 1.1. Overall the system only identified 13 (1.42%) of the 914 judgements made that appeared to be seriously inconsistent (SE exceeded 2).

There was a strong and significant correlation ($r=0.84$ $p<0.01$) between the scores from the comparative pairs judging and those from analytical marking of the portfolios. Similarly there was a strong and significant correlation ($r=0.87$, $p<0.01$) between the rankings from the scores of the comparative pairs judging and the analytical marking. This provides further evidence that assessors may use a form of holistic judgement even for the analytical marking.

7. Assessor Perceptions of Task, Processes and Quality

The assessors were asked for feedback on the suitability of the digital representations of student work, the marking processes and the quality of student work. They were asked a series of questions either in person or via an email. It must be remembered that 10 of the assessors were also teachers of the classes involved in the study. As teachers they had already indicated opposition to digital representation and submission for assessment and therefore were likely to carry that negative perception into the scoring processes. However, these teachers were all experienced in the course and had at least regularly assessed the work of their own teachers.

In general the comparative pairs judging assessors reported that the quality of the digital representations was poor. The photographs were blurry and did not represent the scale, details, textures, media, and dimensions of the real work, especially the 3D works. The videos were reported to be wobbly, shaky, and aside from showing an indication of the size of the artwork, did not contribute much to the assessors' perception of the work. Although some assessors found the video "most insightful" many were highly critical of its effectiveness and lack of "sound and visual quality". One assessor stated that the "voice overs were often uninformative" however another assessor explained that the "video helped to get a better sense of space and form" especially with 3D works. One major problem with the video was that the students were not well prepared to make this presentation and gave minimal information due to the limited time restriction on the video". Many of the videos were "blurry", "wobbly" and the "commentary off-putting ... especially if the student wasn't particularly articulate".

Most assessors felt that the artworks were not well represented by the digitisation. However, a few felt that digitisation adequately represented the artworks, mainly the "traditional paintings". Only one assessor thought that the digitised artworks were "very clear" and they had a "good sense of what the work would be like in reality". The assessors were concerned that intricate features of works were not fully represented such as "fine brushwork, texture and fine skills". Also, "digital images did not project the scale, ambitious composition or visual impact of the artwork". This issue was especially noted for larger artworks including 3D works that require walking around the work. One assessor explained that the "the videos did help in this [issue], to get a better sense of space and form". Some poor quality images as well as a "blurry and shaky video" did not

represent the work well. One assessor explained that to take good photos of the artworks requires a "studio lighting and studio conditions". Even so, a few assessors commented that the digital images were better than for the first phase of the study. It is important to recall that the students took their own photos.

Overall the assessors considered that the digitisations were not an adequate representation of student performance. Some commented that the digitisations "flattened work" and it was "very difficult to see fine detail or subtle use of media/skills and processes". Assessors explained that it was difficult to determine the "scale/magnitude" and "visual impact" of the artwork. One assessor felt that it was difficult for the marker to get an "emotive response" from the artwork by viewing the digitisation. Another assessor commented that it would be useful to "see the preparation of the artwork" and that this was a problem that it no longer was part of the course. A number of assessors stated that they thought the digitisations provided an "indication" of students ability and the quality of the work. Assessors found that the photos and artist statements were the most useful and the video was only somewhat useful.

Nearly all assessors found the judgement making process easy, however, it became more difficult towards the end. This was because the ACJS is adaptive, and as the marking progresses, provides portfolios to compare that are closer in performance standard. However, in another respect the judging became easier as the assessors were familiar with the work they had already viewed and could readily compare it to a new work. An issue which the assessors found that impacted on the judgement making process was the lack of functionality of the system to present two portfolios "side-by-side on one computer screen". One assessor would have liked a "neutral" option, as sometimes "both were equally bad (or some equally good for differing reasons)". The assessors made a number of suggestions to improve the functionality of the ACJS tool and the judging process. These included, designing the system specifically for assessing art, such as improving the viewing functionality of the digital photos (zoom function), and to be able to see the "whole artwork on the screen" including a "360 view of the artwork".

The three analytical assessors were asked to provide feedback on both the analytical and comparative pairs scoring processes. One assessor commented that "both methods are essential in providing a fair and valid judgement of artworks which are inherently subjective". Another assessor stated that the "analytical marking forces the marker to consider all the criteria associated with art making" and it is a "more thorough way to assess a student". One assessor found that the comparative pairs judging was difficult when "making judgements between poor quality 2D and 3D artworks" and dealing with "differences in genre, materials and styles". Some were concerned that comparative pairs judgements could "lead to rash judgments being made" without thorough assessment of the work. There were also concerns that because of the subjectivity of Visual Arts that it is "very difficult to decide which work is 'better'". Assessors that preferred comparative pairs judging explained that it was "easy to choose a good work over a bad one" and that the "comparative process is most effective as judgements are made of one above or below

the other as a standard of measure and ranking". In reference to both analytical and comparative pairs scoring for Visual Arts, one assessor commented:

*Each of the assessment methods has its uses however assessment in the visual arts should never rely **solely** on an assessment of the final product.*

Overall the assessors preferred that the artworks "be best assessed in real life using full objective assessment criteria". This was related to concerns that digitisation for summative assessment might lead to teachers encouraging their "students towards making work that photographs well … this isn't what the course is about but as teachers we are driven by results and the expectations of our schools and parents". In terms of summative assessment, overall the teachers and assessors did not support digitisation. However they were supportive of the use of digitisation and comparative pairs marking for setting and understanding the standards in *Visual Arts.* In regard to understanding standards, two assessors commented:

I found the process very informative in marking my own students art works and understanding the application of art forms in the Stage 2 and 3 courses.

It was an interesting experience and I liked the process of comparing and ranking and defining my judgements, I would participate in this sort of assessment process again.

8. 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 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. Therefore at this stage the results would suggest that the use of a variety of digital photographs, and an artist's statement document, are adequate to represent visual artwork in different media and contexts for the purposes of scoring for summative assessment.

The study found that the best consistency of scoring was provided by the comparative pairs method, probably due to combining the judgements of a larger group of assessors. Therefore online submission and a comparative pairs judgements method of scoring should be considered for portfolio assessment even if only for moderation and standard setting. Many of the teachers saw the benefits of using digitisation and comparative pairs marking to inform their understanding of standards in the Visual Arts course. The use of comparative pairs marking using digitised artworks will help teachers gain a consensus view of the standards required in the course. As a result, a third phase of the study is

underway to investigate whether this can be achieved with a set of teachers in different country towns in W.A. using only online technologies. However, at this stage the results suggest that the comparative pairs method of scoring is better than the traditional analytical method for visual artworks done in a variety of media and contexts and submitted for summative assessment.

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.

Acknowledgments

The theory discussed in this paper and the research upon which it is based are as a result of the work of a research team organised by the Centre for Schooling and Learning Technologies (CSaLT) at Edith Cowan University. The team was led by Paul Newhouse and includes researchers Jeremy Pagram, Lisa Paris, Mark Hackling, Alun Price, Pina Tarricone, Martin Cooper and Alistair Campbell, and a number of research assistants and advisors. The research was conducted with the Schools Curriculum and Standards Authority in Western Australia and partly funded by the Australian Council for Research.

References

Barrett, P. T. (2003). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology, 3*(18), 421–439.

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.

ePhoto Inc. (2012). *ePhoto photography and portable film video lighting kit and support system*. Retrieved January 5, 2013, from http://www.ephotoinc.com/photography-lights-studio-video-lighting-portable-video-lighting-kit.html

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.

Kimbell, R., Wheeler, T., Miller, A., & Pollitt, A. (2007). *e-scape: e-solutions for creative assessment in portfolio environments*. London: Technology Education Research Unit, Goldsmiths College.

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

McGaw, B. (2006). *Assessment to fit for purpose*. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Newhouse, C. P. (2011). Comparative pairs marking supports authentic assessment of practical performance within constructivist learning environments. In R. F. Cavanagh & R. F. Waugh (Eds.), *Applications of Rasch Measurement in Learning Environments Research*. Rotterdam, The Netherlands: Sense Publishers.

Newhouse, C. P. (2014). Using digital representations of practical production work for summative assessment. *Assessment in Education: Principles, Policy & Practice, 21*(2), 205–220.

Newhouse, C. P., & Njiru, J. (2009). Using digital technologies and contemporary psychometrics in the assessment of performance on complex practical tasks. Technology, *Pedagogy and Education, 18*(2), 221–234.

Pollitt, A. (2004). *Let's stop marking exams*. Paper presented at the International Association for Educational Assessment Conference. 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.

Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature Review of E-assessment*. Bristol, UK: NESTA Futurelab.