

2014

Digitizing Practical Production Work For High-stakes Assessments

Christopher P. Newhouse

Edith Cowan University, p.newhouse@ecu.edu.au

Pina Tarricone

Edith Cowan University, p.tarricone@ecu.edu.au

This article was originally published as: Newhouse, C. P., & Tarricone, G. (2014). Digitizing practical production work for high-stakes assessments. *Canadian Journal of Learning and Technology*, 40(2), 1-17. Original article available [here](#)

This Journal Article is posted at Research Online.

<http://ro.ecu.edu.au/ecuworkspost2013/649>

Digitizing practical production work for high-stakes assessments

La numérisation de travaux pratiques de production pour les évaluations à enjeux élevés

C. Paul Newhouse, Edith Cowan University, Western Australia
Pina Tarricone, Edith Cowan University, Western Australia

Abstract

High-stakes external assessment for practical courses is fraught with problems impacting on the manageability, validity and reliability of scoring. Alternative approaches to assessment using digital technologies have the potential to address these problems. This paper describes a study that investigated the use of these technologies to create and submit digital representations of practical production work and forms of creative expression for summative high-stakes assessment. The study set out to determine the feasibility of students creating and submitting these digital representations for assessment and to identify which of analytical or comparative pairs scoring generated the more reliable scores. This paper proposes that scoring digital representations of creative practical work submitted by students is a viable alternative to traditional approaches to assessment.

Key words: Digital portfolio, practical assessment, comparative pairs scoring, analytical scoring

Résumé

L'évaluation externe à enjeux élevés dans les cours pratiques se heurte à des problèmes qui se répercutent sur la gestion, la validité et la fiabilité de la notation. Des approches différentes de l'évaluation utilisant des technologies numériques ont le potentiel de remédier à ces problèmes. Cet article décrit une étude consacrée à l'utilisation de ces technologies pour créer et soumettre des représentations numériques de travaux pratiques de production et de création pour une évaluation sommative à enjeux élevés. L'étude visait à déterminer si la création de ces représentations numériques par les étudiants et leur soumission pour évaluation étaient réalisables. Elle visait aussi à identifier quel système de notation de groupe, analytique ou comparatif, générerait les scores les plus fiables. Cet article soutient que noter les représentations

numériques de travaux pratiques soumis par les étudiants offre un choix viable aux approches traditionnelles d'évaluation.

Introduction

It is readily argued that assessment drives the curriculum and therefore it is important that assessment reflects the intended outcomes or objectives of the curriculum (e.g., Stobart, 2008; McGaw, 2006; Clarke-Midura & Dede, 2010). Assessment is defined as a “judgement which can be justified according to specific weighted set goals, yielding either comparative or numerical ratings” (Taras, 2005, p. 467). The weighted goals described by Taras are matched against specified criteria, which are directly informed by the curriculum. Therefore if it is intended in a course that students learn through practical or creative production work (i.e. the outcome is a tangible artefact as discussed by Dillon and Brown (2006)) then this work should be assessed, because otherwise this work will be devalued by students and teachers (Lin & Dwyer, 2006; McGaw, 2006). For high-stakes external assessment it is important that the form of assessment reflects the requirements of the practical curriculum. However, there are many well-documented obstacles to assessing practical production work including the logistics and costs of managing the processes (e.g., Madeja, 2004) and the difficulty of providing a fair and reliable judgement of the work (e.g., Brookhart, 2013). These obstacles are presented in the *Visual Arts* and *Design* senior secondary courses in Western Australia (W.A.) where bulky and/or delicate work is sent to central examination centres at significant cost, some aspects of the work is not included (e.g., process of producing the art, and the prototype of designed product), and expensive scoring processes (e.g., assessors have to attend a central location) generating scores of questionable reliability. These courses are pre-tertiary typically taken by Year 11 and 12 (age 15-17 years), and thus very high-stakes with outcomes from assessment contributing substantially to University entrance. Typically our societies have addressed problems by developing technologies, so it is reasonable to look for technological solutions to these problems of assessment. Could digital technologies contribute to solutions?

It is of paramount importance for summative assessment that measurement is both reliable and valid (Stobart, 2008). Assessing practical work in these types of courses requires judgements to be comparable between contexts, scores divorced from the subjectivity of assessors, and management of the work of large groups of students spread across wide jurisdictions. For example, in the *Visual Arts* course student practical submissions may include two-dimensional, three-dimensional or digital artefacts. Thus judgements must be comparable between the different media of the artefacts. For the *Design* course the portfolio largely presents the process of design and very little representation of the final product whereas for the *Visual Arts* course the only reference to process is an “artist statement,” which is typically less than one page.

Therefore our study set out to investigate the feasibility of representing the practical work in these two courses in digital forms for the purpose of summative assessment and online scoring using the comparative pairs method. This was the main research question and would have application to many other courses facing similar types of problems. The study focussed on three subsidiary questions (paraphrased): (1) What techniques and procedures are appropriate to represent the work in digital forms? (2) Does the comparative pairs judgements method deliver reliable scores for the digital forms of this work? (3) Are these scores consistent with those from

assessing the physical forms of the work? However, this paper discusses the findings more in terms of the main research question concerning feasibility.

The study used the notion of feasibility developed in the British *e-Scape* project (Kimbell, et al., 2007) and their analytical four-dimensioned framework describing feasibility in terms of manageability, technical affordance, functional operation and pedagogical alignment. This framework is discussed later in the paper and was adopted to allow comparisons to be made with this project upon which the study was built. Technologies exist to create digital representations (e.g., photographs, audiovisual recordings, graphic and text files) of the various forms of physical artefacts, personal expression and documents that are representative of practical work in these courses; however, the question is whether this can be done to authentically represent the work and at a reasonable cost. As Dillon and Brown (2006) argue, the resulting digital portfolio must demonstrate adequate fidelity to gain the confidence of all stakeholders, including assessors, students and teachers. Then there is the problem of reliably scoring this work that relies on human judgement of subjective concepts such as creativity and form of communication. There is evidence that reliability will be enhanced with the input of more expert judges. Thus the study brought together three major areas of knowledge: portfolio assessment, psychometrics, and computer-supported assessment.

Portfolio assessment is commonly used by teachers in many areas of the curriculum and with all ages of students handing in projects or folios of work (e.g. Madeja, 2004). However, it is less used for high-stakes summative assessment because as Koretz (1998) identified they are difficult to mark reliably and are resource intensive and management is “problematic” (p. 309). Portfolios vary in form and purpose, for example, Messick (1994) distinguished between a performance portfolio, which concerns processes and procedures, and a product portfolio; with the extent to which each is included in an assessment depending on how clearly task procedures may be determined and varied. In the *Visual Arts* course the focus of the portfolio assessment was on the product whereas for the *Design* course the focus was on the processes and procedures. As a result of the creative work there is a physical artifact, which is the product of the creative expression.

Psychometrics refers to the measurement of psychological attributes through representing qualitative mental processes numerically (Barrett, 2003). The judgement of practical work associated with *Visual Arts* and *Design* is clearly subjective or qualitative in nature and yet summative assessment requires quantifying this judgement. In a sense traditionally teachers have done this through allocating marks or scores based on their judgements of aspects of the work and most often they have added a series of such scores to produce a total score. A British psychometrician Pollitt (2004) explains how this is likely to be very inaccurate and suggests the use of more holistic approaches. He is a more recent advocate for a comparative method that has a history of nearly a century where assessors progressively select the better performance between pairs. This has been impractical on a large scale until the recent development of online systems to facilitate the processes, such as the system used by the *e-Scape* project (Kimbell, Wheeler, Miller, & Pollitt, 2007). The method relies on the use of Rasch statistical modeling to generate scores on an interval scale, and has delivered highly reliable sets of scores (Pollitt, 2012), including in our own research (Newhouse, 2010).

Computer-supported assessment is a term used to describe a large number of ways in which computer technology may be used in assessment whether that be students using the technology to

complete an assessment task or teachers using the technology to manage or mark the assessment output (Bull & Sharp, 2000). Applications of computer-supported assessment have been developed in response to the shortcomings of traditional forms of assessment that in particular are seen to fail to assess more complex tasks and higher-order thinking skills (Lin & Dwyer, 2006; McGaw, 2006; Pellegrino & Quellmalz, 2011). The use of portfolios has typically been offered as one means of addressing these issues but has presented some obstacles, particularly in terms of manageability and measurement reliability (Clarke-Midura & Dede, 2010). Therefore, 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). Our study focused on digital portfolios to represent particular forms of creative expression for the purpose of summative assessment. Dillon and Brown (2006) have provided a theoretical framework for describing and using these types of portfolios in the creative arts.

This paper discusses the implementation and some of the findings from a three-year study investigating the potential to use digital representations of practical work for summative assessment in two senior secondary courses in W.A.: *Design*; and *Visual Arts*. However, these findings are relevant to assessment practice and policies for any course that includes practical work. The study commenced in 2011 and focussed on how the current practical work in these two courses could best be represented in digital form by students, and the best method of scoring these digital portfolios to generate reliable scores. This built upon the success of five years of research into digital forms of assessment conducted by some of the chief investigators (Newhouse, 2010) and collaboration with researchers in the British *e-Scape* project (Kimbell, et al., 2007).

Method

Our research design was based on pseudo-ethnographic action-research evaluation that involved collecting a range of qualitative and quantitative data in two development-evaluation cycles. The first phase investigated whether digital representations of the practical work could be created with adequate fidelity for the purposes of external scoring. The relative merits of the two methods of scoring were also investigated. The second phase investigated the feasibility of students creating the digital representations in school. This paper mainly focuses on this second phase but necessarily provides some information about the first phase as background. More details on the findings from the first phase may be found in (Newhouse, 2014).

A range of data was collected related to each teacher, student and assessor including: observation of the process of digitizing; a survey of students; interviews with teachers, assessors and groups of students; scores generated by the two methods of scoring; and the official score awarded for the W.A. Certificate of Education (WACE). The WACE scores were not generated by the study but were provided by the awarding body and assessed the physical portfolio. The study used an analytical and a comparative pairs methods of scoring to assess the digital representations of the same portfolios. The analytical method used a standard rubric that described various levels of achievement for a set of criteria and allocated score points to each. The comparative pairs method requires assessors to select the “best” performances from successive pairs of performances based on an agreed holistic criterion, with the results combined using a

dichotomous Rasch model to generate an interval score (refer to Pollitt (2013) for a more detailed description).

Observations of the digitization process were noted and used to inform the overall outcomes of study. In addition, the survey of students and teachers, and the assessor interviews data were used to gauge the perceptions and attitudes towards digitization of practical portfolio work for each school and for all schools involved in the study. The interviews were semi-structured (Patton, 1990) and recorded in digital and note form. All teachers and assessors involved in the project were interviewed and a small group of students from each school who were involved were also interviewed. The surveys were employed to collect data on the attitudes of all students and teachers involved in the study about the digitization of practical portfolio work as well as some background data on their use of digital technologies. These data were analysed to address a four-dimensional feasibility framework adapted from the e-scape project (Table 1).

Table 1:

Descriptions of feasibility framework dimensions (Kimbell, et al., 2007).

Dimension	Description
Manageability	Concerning making the digital representation of a type of portfolio do-able under normal constraints for an examining body.
Technical	Concerning the extent to which technologies can be adapted for use to represent portfolios in digital form.
Functional	Concerning reliability and validity, and the comparability of data with other methods of judging portfolios.
Pedagogic	Concerning the extent to which digitized portfolios align with the preferred pedagogy within the courses.

The first phase of the study used samples of work drawn from both courses that had been submitted for final assessment. Initially the research team conducted a situation analysis that included a review of syllabus requirements and of the nature of portfolios typically submitted, and as a result digitizing specifications were defined for each course and each type of submission (e.g., Visual Arts 2D and 3D). Then the samples of work were digitized by researchers ready to be scored by independent assessors. In the second phase a sample of classes was selected from the two courses and the digitizing specifications were further refined and explained. Researchers then supported the teachers and students in their schools to digitize the students’ own work ready to be scored. Students used specifications and methods of digitization drawn from the results of the first phase of the study.

The *Design* course portfolio comprised up to 15 single-sided A3 paper pages and had to include identification of each project referred to, a designer statement, references and acknowledgements, and the design projects. Students selected examples of design development work from up to three projects to provide “evidence of their understanding of, and practical skills in, the generation and production of design.”

The *Visual Arts* course portfolio was known as the practical submission that had to include: a resolved artwork; a declaration of authenticity; the artist statement; a references/ acknowledgement form; and a photograph of the completed resolved artwork for submission. There were three categories of submission: two-dimensional; three-dimensional; and motion and

time-based. Each category had defined constraints such as for two-dimensional artwork it must not exceed 2.5 sqm and 20 kg. There were no “motion and time-based” artwork in our sample.

Analytical and Comparative Pairs Scoring

Analytical scoring criteria were taken from the course documentation and represented as a rubric. This was the same rubric used for the official marking of the physical portfolios. Marks are allocated for each criterion along with descriptions of the required performances for each score point. Then for each course we developed a holistic criterion based on the analytical criteria, to use for comparative pairs scoring. These were developed collaboratively by the assessors at a workshop. Finally, purpose-built online tools were used by assessors to implement each of the methods of scoring.

FileMaker Pro (Filemaker Inc., 2007) relational database software was used to create an online analytical scoring tool for each course. Assessors used a standard web-browser to logon to a list of student IDs and then for each student a scoring screen with the rubric and displays of the digital representations. Radio-buttons underneath each criterion were used to indicate a mark that was then automatically added to the total.

For comparative pairs scoring the online *Adaptive Comparative Judgements System* (ACJS) developed for the *e-Scape* research project was used. Pollitt (2012) describes in detail how the *ACJS* combines all the main processes involved in scoring using the comparative pairs method including generating the pairs of portfolios for each assessor to judge, provided a facility for recording that judgement and maintaining assessor notes, and calculating scores and associated reliability coefficients using Rasch modelling. This meant that assessors only needed to judge pairs until an acceptable level of reliability was attained. For each course an initial half-day workshop was conducted during which the first rounds of judgements were completed. Assessors were then able to continue the process from home or work over a four-week period. We had predetermined to stop when the Cronbach’s Alpha reliability coefficient became 0.95 and this happened during the 13th round for both courses in the first phase of the study.

Digitization of Practical Work

For the first phase of the study the samples of the practical submissions digitized by researchers were associated with 75 *Visual Arts* students and their 10 teachers, and 82 *Design* students and their 6 teachers. The submissions were each represented in one or more digital files of various types that were stored on servers as digital repositories. These digital representations were scored by external assessors using the online tools for the two methods of scoring: analytical marking by two or three expert assessors; and comparative pairs judging by some of the teachers and other expert assessors.

For the second phase of the study Year 11 students from 13 *Visual Arts* schools (138 students with 105 female) and 14 *Design* schools (110 students with 59 male) digitized their own practical work based on the technical specifications and submitted it to the MAPS online system for external scoring. Teachers and students were supported during the digitization process with guidance and the provision of equipment. Two or three school visits of up to 2 hours were scheduled with the first designed to outline the technical specifications and the process of digitization, and the others to facilitate the digitization process.

Surveys and Interviews

The student survey was administered after submission of digitized portfolios, for both courses and both phases. The questionnaire consisted of 25 closed response items and four open-response items. Broadly, it sought students' experience with, and opinions on, assessing their practical work using digital representations, use of computers and other digital devices, attitudes to using computers, and facility with computer applications. Three scales were derived from combining items from the questionnaire. Descriptions of these scales are as follows.

<i>eAssess</i>	A measure of the efficacy of the digitization based on responses to a set of items with four response options, resulting in a score between 1 and 4.
<i>Skills</i>	A self-assessment measure of ICT skills based on responses to a set of items related to software/hardware applications and with four response options resulting in a score between 1 and 4.
<i>SCUse</i>	Estimate of time in minutes per day spent using computers at school based on responses to items asking student to recall time spent on each day of the previous week.

The open-ended questions asked students to list the two best things and two worst things about their digital portfolio, what additional technology they would have liked to use, and the types of software and technology they used to create their portfolio.

The main findings are now discussed for each course separately followed by a combined summary of outcomes for practice and policy organised using the feasibility framework. It is not intended in this paper to provide a full report of the results of data analysis, rather relevant summaries or statistics are reported to support findings. For example, reliability statistics are quoted to support findings related to the functional operation of the assessments but descriptive statistics on the scores are not because they have little relevance to the findings that are discussed. More details on the results of scoring in the first phase may be found in (Newhouse, 2014). The main focus is on the second phase of the study and the perceptions of students, teachers and assessors

Findings for Design Portfolios

For the first phase of the study our sample of 82 portfolios were intercepted at a central location where all portfolios had been sent in school bundles from across the state. An automatic colour A3 scanner was used to convert each portfolio into a single PDF file. These files were uploaded to the online repository for use in analytical scoring and copied to the *ACJS* server for comparative pairs scoring. The two analytical assessors working independently took a total of about 17.5 hours, and generated scores that were only moderately correlated ($r=0.53$, $p<0.01$), representing a low inter-rater reliability. Further, Rasch analysis revealed inconsistency for some of the criteria that tended to discriminate poorly and be scored too high. For comparative pairs scoring the *ACJS* system allowed assessors to view both portfolios side-by-side or separately and the images of the pages could be expanded for viewing. The system estimated a total time of a little over 50 hours. Correlation analysis between the three sets of scores found moderate coefficients between each with the strongest between the *Pairs* scores of the digitized portfolios and the official WACE scores of the paper-based portfolios (Newhouse, 2014).

The assessors generally believed that the scanned files represented the work well but some suggested a higher resolution was required. Most considered that the comparative pairs method of scoring would increase the reliability partly because they believed that having only one holistic criterion eliminated the likelihood of different interpretations. The teachers and most students were generally positive toward the concept of digitizing the design portfolios for assessment but only if this was done by the students. All the teachers had their students working in digital modes for logistical reasons and because they saw as important to “keep pace with industry.” Students tended to be disappointed with the quality of the scanned portfolios and felt that submitting their own digital portfolio would allow them to present their work more professionally, improve the editing of their work and give them the opportunity to analyse their work critically. They indicated some previous experience in doing so and that they had the necessary skills with computers, cameras and editing software.

From the first phase it was clear that although the scanning of the paper portfolios and the online scoring was successfully accomplished, the outcomes were not ideal (Newhouse, 2014). It was likely that the quality of portfolios would be improved if students created and submitted them in digital form as a result of the increased flexibility and the opportunity for including audiovisual content in addition to the text and graphics. The poor consistency between assessors was related to the structure of the portfolios and marking criteria. The portfolios needed to be less complex and contain less information and the analytical criteria needed to be more explicit for each score point. However, these could not be substantially modified for the second phase as the portfolio had to match the course requirements.

The specifications for the second phase required that the students create one PDF file comprising 15 screens but also to create a 20 second video focusing on an aspect of their work they wanted to articulate for the markers. Students were provided with digital video cameras and computers with Adobe Acrobat Professional software. In preparation they were asked to locate their electronic and hard copy files for the portfolio; however, few managed their electronic files well and therefore many spent considerable time preparing for digitization. Students did not necessarily perceive file management as difficult but did not value it as a requirement for assessment. Very few students had completed all work electronically and so they scanned hard copy pages into PDF files using a school photocopier/scanner and then these files were inserted into the sequence of pages for their digital portfolio. A problem identified by the students was that often pencil sketches did not scan very well and therefore were difficult to read. In general they were able to create the combined PDF files with minimal direction and in most schools upload the files into MAPS. For some files a small amount of editing was required in preparation for scoring (mainly rotating of pages).

Table 2 presents the frequency of student responses to some of the questionnaire items related to the creation of the digital portfolios and how well it represented their work. Although 47% of students created “lots” of their digital portfolio digitally, 38% of students responded that “none” of their digital work was marked electronically. This would be mainly due to current paper-based portfolio submission requirements for the external examination. Availability of technology in some schools was also an issue, especially with the types of software that students were exposed to and encouraged to use to develop their portfolio. The vast majority of students were confident in creating their digital portfolio and also considered that it represented their work very well. They also found the process of digitization easy and the technical specifications easy to follow.

Although most of the portfolio work was completed digitally, about half of the students preferred to submit a paper portfolio for assessment; this may have been due to paper submission being current practice.

Table 2:

Frequency of responses to questionnaire items about the Design digital portfolios for the second phase of the study.

Items	Frequency of responses (%)			
	Lots	Some	Little	None
How much of your Design work is created digitally?	47	36	13	0
How much of your work has been marked in digital form before?	9	33	17	38
How much of your PDF submission was originally created on paper?	28	28	38	6
How much help did you need to digitize your portfolio?	11	36	38	14
	SA	Agree	Disagree	SD
The PDF and video represent my work very well.	5	70	21	3
I felt confident in creating my PDF and video.	12	62	21	4
I prefer to submit my portfolio in paper form.	28	36	27	7
I liked being able to include a video with my digital portfolio to explain my design.	9	31	47	11
I would prefer an assessor to mark a paper version rather than the digital version.	15	37	43	2
I think I would get a higher mark for the paper version rather than the digital version.	12	44	40	7
It was easy to combine the elements of my portfolio into the PDF file.	21	57	16	5
The technical specifications for my digital portfolio were simple to follow.	13	71	12	3

Note: the rows do not necessarily sum to 100% because some students did not respond to all items

Some descriptive statistics for the three scales derived from the questionnaire are shown in Table 3. The *eAssess* scale had a mean around the mid-point of 2.5 indicating that generally the students considered that the digitization process was effective in capturing and representing their Design portfolio work. However, scores went as low as 1.5 indicating a level of dissatisfaction among many. The mean for the *Skills* scale was well above the mid-point indicating that generally these students perceived that they had a high level of skills with the use of computers and other digital devices. The third scale, *SCU*se, indicates that these students’ estimated computer usage at school was, on average, nearly two hours a day that would be regarded as relatively high.

In open-response items many students stated that the improved appearance of graphics including colour was a positive aspect of their digital portfolio, that the work looked neater, and the digitization process was fun, practical and easy. However, some found the video was “extra work” and “awkward” to create, and that the digitization process was “time consuming.”

Table 3:
Descriptive statistics for the scales based on items from the Design student questionnaire for the second phase of the study.

	N	Min	Max	Mean	SD	α	Description
eAssess	105	1.5	3.6	2.4	0.4	0.75	Efficacy of the digitization. Score between 1 and 4
Skills	106	1.5	4.0	3.1	0.5	0.73	Self assessment of ICT skills. Score between 1 and 4.
SCUse	107	0	480	112	108	n/a	Estimate of time in mins/day using computers at school.

α = Cronbach’s Alpha reliability coefficient

Teachers were interviewed and asked to provide feedback on the digitization of the Design portfolios. They were overwhelmingly supportive of the submission of digital portfolios for assessment, for example stating that it is “a lot easier to submit portfolios this way” and that it was a “sensible solution.” Many of them commented that it was important that the portfolio “looks professional” and reflects industry requirements and standards. Teachers from country schools thought that the digital portfolio was a “better way for regional schools to submit work and was better for the student.” Importantly, the teachers felt that the digital portfolios were “preparing our students for the real world” and that the “techniques are invaluable for future courses in design.” Some negatives related mainly to the need for students to “improve file management skills [so they] have a more organised approach to creating pages, edit them and get them in the correct order.” Teachers commented that the technical specifications worked well, but the students needed to be better prepared at the beginning of developing a portfolio for the digital submission requirements. Some teachers and students, particularly from country schools, mentioned technical problems encountered in uploading files into the MAPS system.

Findings for Visual Arts Portfolios

For the first phase of the study teachers had organised for their students’ artwork submissions to be delivered to a central location that in some cases was over 1,000 kilometres. Our sample of 75 submissions was photographed and video-recorded on one day by three teams of researchers using SLR digital cameras, digital video cameras, and a motorised turntable for 3D work. The resulting files were uploaded to the digital repositories so that each student had main artwork photos, four digitally created close-up photos, a photo of the artist statement, a video, and a PDF containing all the photos. The three analytical assessors working independently took a total of just over 37 hours, and generated scores that were only moderately correlated ($r=0.51$ to 0.56 , $p<0.01$), representing a low inter-rater reliability. Rasch analysis revealed that overall the criteria tended to discriminate reasonably well providing a good reliability for the measure. For comparative pairs scoring the ACJS online tool presented each portfolio in eight boxes but only one portfolio could be viewed at a time. Clicking on each box would open a screen displaying the contents clearly. The system estimated a total time of a little under 45 hours. Correlation analysis between the three sets of scores found moderate to high coefficients between each. These correlations were much stronger than between the scores from the three analytical assessors perhaps representing that their combined judgement and that of the 14 comparative pairs assessors was superior to their individual subjective judgements (Newhouse, 2014).

The assessors perceived that the digital representations were poor due to the quality of the photographs and videos and that multiple-piece works lacked coherence. They believed that the scale and critical features of the work were not adequately represented, and the videos contributed little. All of the teachers and most students were opposed to the idea of using digital representations for assessment. Some students felt their work was not adequately represent their artwork; however, most indicated that the photographing was done quite well. Almost half would have preferred to create the digital portfolio themselves although most indicated little experience in representing their artwork digitally (44% no experience) and a sizeable group indicated a low level of skills particularly for video editing. Teachers were concerned about inequities and the potential for fraud if the digitizing was done at school and four doubted the technical capacity of their students.

From the first phase it was clear that although the digitizing of the practical submissions and the online scoring was successfully accomplished this was difficult and time-consuming and not scalable where digitization occurs at a central location. Even so, there was little evidence that the quality of digitization influenced the outcome of scoring the work (Newhouse, 2014). Although they would need support there was evidence that students would be able to represent each type of artwork in digital forms.

For the second phase of the study students were required to take a specified number of digital photos and also a 20 second video of their artwork, as outlined in the technical specifications. They were provide with SLR digital and video cameras along with inexpensive portable backdrops and lighting. They were instructed to have a 2D or 3D artwork ready and they were guided on how to use the equipment prior to digitization. Apart from the technology support and the technical specification requirements, the students were autonomous in making decisions about what they wanted to emphasise to markers in their artwork through the close-up photos and the video. The video provided an opportunity to explain their work whilst zooming-in on chosen aspects for emphasis.

Table 4 provides the frequency of student responses to questionnaire items related to the creation and efficacy of the digital photographs and video of their artwork. The majority had previously used very little technology to create or capture their artworks. This could be due to there being little requirement to use technology in their curriculum and also the requirement of submitting the original artwork for the external examination. However, the majority felt that they required very little help to digitize their artworks. Overall the students were split between being confident and not confident in creating their digital representations. Although they generally found it easy to create their digital representations, over half would have preferred others to carry out the digitization. Most agreed that the digital photographs and video represented their artwork very well. Most (90%) preferred the original artwork to be marked and 80% considered that they would obtain a higher mark for the original artwork than the digital representation.

Descriptive analysis for the three scales derived from the questionnaire is shown in Table 5. The mean for the *eAssess* scale was above the mid-point of 2.5 indicating that generally the students considered that digitization process was an effective means of representing their artwork. However, as for the Design students some scores were much lower indicating a level of dissatisfaction. The mean for the *Skills* scale was well above the mid-point although not as high as for the Design students. In general these students perceived themselves to have good skills with the use of computers and other digital devices. However, they only spent on average 60

minutes a day using computer technology at school, as indicated by the *SCUSe* scale mean, well below the mean for the Design students.

Table 4:
Frequency of responses to questionnaire items about creating Visual Arts portfolios for the second phase of the study.

Questionnaire items	Frequency of responses (%)			
	Lots	Some	Little	None
How much of your visual art work has been represented on a computer before?	6	15	31	45
How many visual art works have you submitted in digital form for marking?	1	5	12	79
How much of your submission was originally created on computer?	3	8	19	67
How much help did you need to digitize your visual art work?	12	29	38	17
	SA	Agree	Disagree	SD
The digital photographs and videos represent my Visual Arts work very well.	1	61	26	5
I felt confident creating the digital representations myself.	4	50	33	9
I would prefer someone else to create the digital representation.	18	37	37	4
I would like to include additional information with my digital representations.	7	54	33	1
I would prefer an assessor to mark my original artwork rather than the digital.	49	38	10	0
I think I would get a higher mark for my original artwork than the digital.	29	51	16	0
It was easy to create my digital representations	11	59	23	3
The requirements for my digital representations were simple to follow	16	65	13	1

Note: the rows do not necessarily sum to 100% because some students did not respond to all items

Table 5:
Descriptive statistics for the scales based on items from the Visual Arts student questionnaire for the second phase of the study.

	N	Min	Max	Mean	SD	α	Description
eAssess	126	1.6	3.5	2.6	0.3	0.65	Efficacy of the digitization. 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’s Alpha reliability coefficient

In open-response items many students felt that it was good to have a photo of the whole work and that it was “easy,” “fun” and “good to have a digital copy.” They also considered that the close-ups showed the detail in their work and “emphasised tonal modulation.” They felt that digitization provided “quality clear images” and “good light, shadow and colour capture,” although some considered that digital images “hide faults” and “can make [the] artwork look better.” Some felt that the digitized work “wasn’t necessarily a true or good representation” of the original work, didn’t look as “impressive in digital form,” and did not show “fine details, texture and doesn’t capture the essence” of the original artwork. However, they felt that the

digital representation would be much “easier” to mark and the logistics supporting scoring would be better. Some commented that the annotated video would be useful for markers to know their “reasons/points of view/hidden messages” about their work.

Teachers were interviewed and asked to provide feedback on the digitization of the student practical work. They teachers made a number of positive comments regarding the outcomes of the digitization including that their students were “capable” of digitizing artwork as long as they were provided with the equipment, training and teacher support. Some considered that scoring the digitized artwork wouldn’t be an “issue” and that “the marks will not vary enough to be too much of a problem.” Although most thought that the digital photos were “clear,” some also felt that they did not capture the “subtle nuances” and “textures” in the artworks. Generally, the teachers considered that the “original work would score higher” and that it was “very risky” to rely on the digital representations because the viewer couldn’t “interact” with the work or appreciate the “scale and subtleties” of the work.

Summary of Outcomes for Practice and Policy

Some initial outcomes for practice and policy are discussed using the dimensions of the feasibility framework adapted for the study and central to the main research question.

Manageability

For *Design* managing the scanning of the paper-based portfolios and converting them to PDF files was very straight forward and the resulting files were readily made available to assessors for both methods of scoring. With a minimum of guidance students were confident and found it relatively easy to follow technical specifications to digitize their own work. Some needed to have better file management.

For *Visual Arts* creating the digital representations of the artworks at a central location was difficult and time-consuming and not scalable. However, the process was much more efficient and the outcomes were much better when the students digitized their own work, including setting up backdrops, lighting and cameras. They needed to be prepared in advance to video and annotate their artwork. The digital representations were readily made available to assessors for both methods of scoring. Some teachers did see some management and cost advantages in digital representation due to not having to transport the work.

Technical feasibility

For *Design* the colour A3 scanner readily converted paper portfolios into single PDF files. Very little editing of files was required for scoring. However, most teachers felt the quality of the scans was not adequate; lacking in resolution and accurate colour reproduction. The video camera and online technologies used by students worked well and most schools had the software required to create the digital portfolios. However, the quality of the digital portfolios was likely to be affected by the design capability of the software available in each school, and the need to scan some work done on paper.

For *Visual Arts* it was demonstrated that it was technically possible to adequately represent each type of artwork in digital forms using SLR digital and video cameras and the specifications developed for the study. A small amount of reformatting of files was required in preparation for online scoring. Most students were capable of using the equipment to digitize their own artwork

but needed instruction in the skills and requirements necessary to do so effectively. There were concerns about the resolution of images, however, this did not appear to noticeably influence the results of scoring.

Functional feasibility

From the first phase of the study for *Design* there was poor consistency between assessors for analytical scoring although the combined scores exhibited reasonable reliability (Newhouse, 2014). There was a high level of reliability for the results of comparative pairs scoring. However, when comparing the scores for the digital and paper representations, and between the different methods of scoring, there were only low to moderate correlation coefficients. It is likely that this was due to the quantity and complexity of information in the portfolios making it difficult to judge using a set of criteria. Most teachers and students indicated that using digital representations of design work was a more valid method of assessment than the existing paper-based portfolios provided that the students created and submitted their own.

From the first phase of the study for *Visual Arts* there was also poor consistency between assessors for analytical scoring probably due to a high level of subjectivity in the perception of the artworks (Newhouse, 2014). There was evidence that the most reliable measure would be generated by the comparative pairs method of scoring with the combined judgement of assessors appearing to negate a lot of this subjectivity. There was a high level of correlation between scores for the digital representations and the original artworks that had been through a rigorous moderation process. However, most students and teachers would not like artwork to be assessed using digital representations, and felt that the only valid approach was to judge the original artworks.

Pedagogic feasibility

For *Design* most students already completed at least some parts of their portfolios digitally and therefore they, and their teachers, believed that submitting digital portfolios was better aligned with the intentions of the course. Teachers saw advantages including future use for job applications but about one-third of the students indicated limited experience and felt they would need some time to become proficient. Students tended to find the digitization process easier to conceptualize and complete if they had been given more experience with better quality software in the course.

For *Visual Arts* some students and teachers perceived value in representing art digitally but not for assessment purposes. Some teachers would not be comfortable with students digitizing work for assessment because the quality would be inconsistent and that subtle features of artworks would not be captured. Despite this they tended to agree that the technical specifications provided clear guidance helping to ensure consistency in the quality of the digitized artworks.

Conclusion

Overall the study has found that that it is feasible to use inexpensive digital technologies to create digital representations of practical work in the *Design* and *Visual Arts* courses adequate for the purposes of summative assessment. Further, it is feasible to use relatively standard and accessible online systems to support the scoring of this work with fairly minimal maintenance requirements. This allows the comparative pairs method of scoring to be employed that appeared

to provide reliable scores for both courses and, in particular, appeared to be better suited to the *Visual Art* work, than analytical methods of scoring.

The attitudes of students and teachers towards the use of digital portfolios for summative assessment was quite different for the two courses, very positive for *Design*, particularly if the students created them, and generally negative for *Visual Arts*. The latter did see the value of digitizing to support the learning process and for archiving purposes but ultimately believed that the original artwork needed to be viewed by the assessor. For *Design* the initial limitation to static text and graphics meant that there was little difference between the scanned digital and paper representations whereas the inclusion of non-static media such as the video provided more flexible communication. 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. For both courses external digitization was too cumbersome, time consuming and labour intensive, but it was feasible for students to digitize their own work for submission through an online portfolio system. For these courses and others with similar types of practical work, our study has indicated the feasibility of assessing digital representations, particularly using the comparative pairs method of scoring, but much still needs to be done to gain the acceptance of some of the main stakeholders including students, teachers and the community (Stobart & Eggen, 2012).

Acknowledgements:

The study discussed in this paper was the work of a research team led by Paul Newhouse and included researchers Jeremy Pagram, Lisa Paris, Mark Hackling, Martin Cooper, Pina Tarricone, Alistair Campbell, Alun Price and many research assistants. The work of everyone in this team, particularly Martin Cooper and Pina Tarricone, and the teachers and students involved, contributed to the research outcomes presented in this paper.

References

- Barrett, P. T. (2003). Beyond Psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 3(18), 421-439.
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69-90.
- Bull, J., & Sharp, D. (2000). *Developments in Computer-Assisted Assessment in UK Higher Education*. In R. Sims, M. O'Reilly & S. Sawkins (Eds.) conference proceedings, Learning to Choose: Choosing to Learn, Queensland, Australia, pp. 255-260.
- 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.
- Filemaker Inc. (2007). *Filemaker Pro 9*. Santa Clara, CA: Filemaker, Inc. Retrieved from <http://www.filemaker.com/products/filemaker.html>
- 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.
- Koretz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education*, 5(3), 309-334.
- 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.
- 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. Retrieved from http://www.iaea.info/documents/paper_1162a2541.pdf
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Newhouse, C. P. (2010). Aligning Assessment with Curriculum and Pedagogy in Applied Information Technology. *Australian Educational Computing*, 24(2), 2-5.
- Newhouse, C. P. (2014). Using digital representations of practical production work for summative assessment. *Assessment in Education: Principles, Policy & Practice*, 21(2), 205-220.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbery Park, CA: SAGE Publications, Inc.
- Pellegrino, J. W., & Quellmalz, E. S. (2011). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education*, 43(2).
- Pollitt, A. (2004, June). *Let's stop marking exams*. Paper presented at the International Association for Educational Assessment Conference, Philadelphia. Retrieved from http://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. (2008). *Testing Times, The Uses and Abuses of Assessment*. Abingdon: Routledge.

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

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466-478.

Authors

C. Paul Newhouse. Email: p.newhouse@ecu.edu.au. Paul is an Associate Professor at Edith Cowan University in Western Australia where he is the director of the Centre for Schooling and Learning Technologies (CSaLT) in the School of Education. His aim is to improve the opportunities for children to develop their potential through engaging and relevant schooling.

Pina Tarricone. Email: p.tarricone@ecu.edu.au. Pina's PhD in educational/cognitive psychology won the 2007 Edith Cowan University Research Medal. Psychology Press published her thesis research as a book titled *The Taxonomy of Metacognition*. She has a M Ed in Interactive Multimedia and has just completed a M Ed in Educational Measurement. Her interests include educational psychology constructs, psychometrics and the use of technologies for assessment.



This work is licensed under a Creative Commons Attribution 3.0 License.