Ergonomics
Vol. 51, No. 9, September 2008, 1352–1375

Taylor & Francis
Taylor & Francis Group

Computer- vs. paper-based tasks: Are they equivalent?

Jan M. Noyes[a]* and Kate J. Garland[b]

[a]*University of Bristol, Department of Experimental Psychology, 12A Priory Road, Bristol BS8 1TU, UK;* [b]*University of Leicester, School of Psychology, Henry Wellcome Building, Lancaster Road, Leicester LE1 9HN, UK*

In 1992, Dillon published his critical review of the empirical literature on reading from paper vs. screen. However, the debate concerning the equivalence of computer- and paper-based tasks continues, especially with the growing interest in online assessment. The current paper reviews the literature over the last 15 years and contrasts the results of these more recent studies with Dillon's findings. It is concluded that total equivalence is not possible to achieve, although developments in computer technology, more sophisticated comparative measures and more positive user attitudes have resulted in a continuing move towards achieving this goal. Many paper-based tasks used for assessment or evaluation have been transferred directly onto computers with little regard for any implications. This paper considers equivalence issues between the media by reviewing performance measures. While equivalence seems impossible, the importance of any differences appears specific to the task and required outcomes.

Keywords: computer vs. paper; NASA-TLX workload measure; online assessment; performance indices

1. Introduction

The use of computer in comparison to paper continues to attract research interest. This is not necessarily in terms of which medium will dominate, although there are still publications on the 'myth of the paperless office' (see Sellen and Harper 2002), but rather on the extent of their equivalence. Testing, for example, is central to the disciplines of Applied Psychology and Education and, in situations requiring assessment, online administration is increasingly being used (Hargreaves *et al.* 2004). It is therefore important to know if computer-based tasks are equivalent to paper-based ones and what factors influence the use of these two media. The aims of the present paper are twofold: (1) to provide a critical review of the more recent literature in this area; (2) to draw some conclusions with regard to the equivalence of computer- and paper-based tasks.

2. Early studies

Experimental comparisons of computer- and paper-based tasks have a long history dating back some decades. Dillon (1992) in his seminal text, 'Reading from paper versus

*Corresponding author. Email: j.noyes@bristol.ac.uk

ISSN 0014-0139 print/ISSN 1366-5847 online

DOI: 10.1080/00140130802170387
http://www.informaworld.com

screens: A critical review of the empirical literature', provided a detailed and comprehensive comparative review. A summary of Dillon's and other earlier findings will be given, although it is not the intention to replicate this review here. It is evident that earlier comparisons focused on traditional outcome measures, for example, reading speed, accuracy and comprehension. A list of pre-1992 studies is given in Table 1; these are now briefly reviewed in terms of outcome measures.

2.1. Reading speed

In a review of research findings, Dillon (1994) suggested that reading was some 20 to 30% slower (in terms of proof-reading performance) from a computer screen than from paper. Many studies supported this conclusion (e.g. Wright and Lickorish 1983, Gould and Grischkowsky 1984, Belmore 1985, Gould *et al.* 1987a,b, Wilkinson and Robinshaw 1987). However, some studies found minimal differences (Kak 1981, Switchenko 1984), while Askwall (1985), Creed *et al.* (1987), Cushman (1986), Keenan (1984), Muter and Maurutto (1991) and Oborne and Holton (1988) reported no significant difference between the two media. Two of the early studies considered a television screen (Muter *et al.* 1982) and video (Kruk and Muter 1984). Both these studies found that reading from the electronic medium was slower.

2.2. Reading accuracy

When considering reading accuracy, findings generally favoured paper. Muter *et al.* (1982), Creed *et al.* (1987) and Wilkinson and Robinshaw (1987) found the degree of accuracy in proof-reading tasks to be lower for computer-based text. However, Askwall (1985), Gould *et al.* (1987a) and Oborne and Holton (1988) reported no significant difference between the two media for accuracy.

2.3. Comprehension

As well as reading speed and accuracy, comprehension had also been studied. Belmore (1985), for example, concluded that information presented on video display terminals (VDTs) resulted in a poorer understanding by participants than information presented on paper. However, there was a caveat to this finding in that it only occurred when the material was presented to the participants on computer first. It appeared that attempting the paper-based task first facilitated using the computer, but not the other way around. This may be partly explained by the fact that participants were not given any practice trials. Belmore suggested that if people had had sufficient practice on the task, comprehension levels should be similar across the two media. There has generally been support for this suggestion since often there is little difference between the attained levels of comprehension (see Muter *et al.* 1982, Cushman 1986, Oborne and Holton 1988, Muter and Maurutto 1991).

Taking a broader definition of comprehension, two studies looked at reasoning (Askwall 1985) and problem-solving (Weldon *et al.* 1985). Askwall showed that there were differences in information searching between the two media with participants searching twice as much information in the paper-based condition (and understandably taking longer), while Weldon found that the problem was solved faster with the paper condition. In terms of output, Gould (1981) found that expert writers required 50% more time to compose on a computer than on paper. Hansen *et al.* (1978) showed that student

Table 1. Studies comparing computer- and paper-based tasks, 1981–1992.

Study	Comparison	Task	Design	Participants	Key Findings
Kak 1981	CRT-displayed vs. printed text	Scanning and comprehension	Within-Ps	4	Significant difference (increase) found in scanning and reading times, and accuracy with CRTs.
Muter *et al*. 1982	Television screen vs. books	Reading continuous text for 2 h	Between-Ps	32	Reading from the television was 28.5% slower.
Wright and Lickorish 1983	CRT vs. paper	Proof-reading	Within-Ps	32	Proof-reading was 30–40% slower with the CRTs.
Gould and Grischkowsky 1984	CRT computer terminal vs. hard copy	Proof-reading for six 45 min periods over 1 d	Within-Ps	24	Proof-reading was 20–30% faster on hard copy.
Keenan 1984	Printed material vs. paper	Four reading tasks: recognition; sentence verification; syntactic editing; misspellings.	Between-Ps	24	Mode had no significant effect on reading rate or accuracy.
Kruk and Muter 1984	Video vs. book	Reading task	Within-Ps	24	Text from video was read significantly slower than from books (214 vs. 237 words per min).
Switchenko 1984	Reading from CRT and paper	Comprehension task following hard and easy article	Within-Ps	Not given	No difference for easy article, but hard article took significantly longer to read on CRT.
Askwall 1985	Computer-supported reading vs. reading text on paper	Reasoning task	Within-Ps	16	No differences found in reading times and accuracy, but difference in information searching.
Belmore 1985	Computer-displayed vs. paper text	Reading four short passages followed by comprehension	Within-Ps	20	Reading significantly slower and comprehension less on computer but effect disappeared when paper presentation first.
Heppner *et al*. 1985	Computer display vs. reading from print	Nelson-Denny Reading Test	Within-Ps	85	Reading performance scores were significantly better on the print forms.

(*continued*)

Table 1. (*Continued*).

Study	Comparison	Task	Design	Participants	Key Findings
Lukin *et al.* 1985	Computerised testing vs. pencil-and-paper	Three personality assessments	Between-Ps	66	No significant difference found in personality assessment scores. Computerised administration preferred by 85% of participants.
Weldon *et al.* 1985	Computer vs. paper manuals	Solving of a problem	Between-Ps	40	Problem solved faster with the paper manual.
Cushman 1986	Microfiche vs. VDT vs. printed page	Reading continuous text for 80 min followed by comprehension	Within-Ps	16	Visual fatigue was significantly greater when reading from negative microfiche (light characters, dark background) and positive VDT. Reading speeds slower for negative conditions; comprehension scores similar across all conditions.
Creed *et al.* 1987	VDU vs. paper vs. photograph of a VDU	Proof-reading	Within-Ps	30 (Expt. 1) 24 (Expt. 2)	1. No significant difference found in time, but significant difference in number of errors for VDU and paper. 2. Two-column format used and results supported those of Expt. 1. Overall, showed that proof-reading was significantly worse on a VDU than on paper.
Gould *et al.* 1987a	CRT vs. paper-horizontal vs. paper-vertical	Proof-reading for mis-spelled words for 10 min	Within-Ps	12 (Expt. 2) 18 (Expt. 3)	1. Significantly faster on paper than CRT. No difference in accuracy. 2. Similar result found for comprehension task.

(*continued*)

Table 1. (*Continued*).

Study	Comparison	Task	Design	Participants	Key Findings
Gould *et al.* 1987b	CRT displays vs. paper	Proof-reading	Within-Ps	18 (Expt. 1) 16 (Expt. 2) 12 (Expt. 3) 15 (Expt. 4) 15 (Expt. 5) 18 (Expt. 6)	1. Reading significantly faster from paper than CRTs. 2. No difference when high-quality monochrome screen was used. 3. No difference with a different regeneration rate. 4. No difference for inverse video presentation. 5. Reading significantly faster from paper when anti-aliasing was taken into account. 6. Five different displays assessed, and no paper.
Wilkinson and Robinshaw 1987	VDU vs. paper text	Proof-reading for four 50 minutes sessions	Between-Ps	24	VDU was significantly slower and less accurate, and more tiring.
Oborne and Holton 1988	Screen vs. paper	Comprehension task	Within-Ps	16	No significant differences shown in reading speed or comprehension.
Gray *et al.* 1991	Dynamic (electronic) vs. paper text	Information retrieval tasks	Between-Ps	80	Dynamic text was significantly better than paper text when answering difficult questions.
Muter and Maurutto 1991	High-quality CRTs vs. books	Comprehension task	Within-Ps	24 (Expt. 1) 18 (Expt. 2)	Skimming was 41% slower with the CRTs. Reading speed and comprehension were equivalent for the two media.

CRT = cathode ray tube; VDT = video display terminal; VDU = video display unit.

participants took significantly longer to answer questions online than on paper. These differences were attributed to the poor design of the interface. Gray *et al.* (1991) attempted to address this. Working on the rationale that electronic retrieval takes longer than paper because of the time taken to access the system and slower reading speeds, they replaced non-linear text with dynamic text. Although practice effects were evident, Gray *et al.* found that information searching improved.

2.4. Summary of findings of early studies

These early comparisons of computer- and paper-based tasks generally favoured paper for better performance according to the metrics of speed, accuracy and comprehension. However, inconsistencies in earlier findings could largely be attributed to variations in visual quality of the two presentations, in that like was rarely being compared with like.

A comprehensive review by Ziefle (1998) reached the conclusion that paper is superior to computer, because of the display screen qualities whereby the eyes tire more quickly. However, comparisons were based on studies conducted in the 1980s and early 1990s. Display screen technology has advanced since this time; therefore, more recent studies should provide findings that are more valid today. Developments in display screen and printing technologies should have reduced the level of disparity between the presentation qualities of the two media and this should be reflected in an improvement in the consistency of findings. Hence, there has been a move away from the traditional indicators shown in the post-1992 studies.

3. Post-1992 studies

More recent studies have increasingly used the traditional indicators in conjunction with more sophisticated measures (see Table 2). Further, there is now a greater awareness of the need for equivalence to be determined fully to ensure that overall performance outcomes are matched; this is especially the case where any decrement may have efficiency or safety implications. This has resulted in many papers specifically comparing computer- and paper-based complete tasks rather than using a partial performance indicator such as reading speed.

3.1. Traditional indicators

Studies using more modern display technology have, somewhat surprisingly, still shown inconsistencies. For example, Mayes *et al.* (2001) found computer-based reading to be significantly slower. Examinations of learning or comprehension, measured in terms of correct answers, have tended not to find differences between materials presented in the two forms (e.g. Mason *et al.* 2001, Mayes *et al.* 2001, Noyes and Garland 2003, van de Velde and von Grünau 2003, Bodmann and Robinson 2004, Garland and Noyes 2004). It should be noted that van de Velde and von Grünau (2003) also found no difference in eye-movement patterns. Rice (1994) also found no difference for comprehension in his first experiment but, when constructs in reading comprehension were examined in a second experiment, a difference was found between the two media in the highlighting of text. In their study, Wästlund *et al.* (2005) found that comprehension from paper was superior (in terms of quantity not quality). Direct comparisons of the two forms of presentation in terms of reading speed and accuracy do not now appear to be of interest, with researchers

Table 2. Summary of post-1992 studies comparing computer- and paper-based tasks.

Study	Comparison	Task	Design	Participants	Key Findings
Horton and Lovitt 1994	Computer (Apple II E) vs. paper	Reading inventories	Between-Ps	72	Group analyses significantly favoured the computer.
Oliver 1994	Proof-reading on screens and paper	Text of around 2000 words	Within-Ps (but all did the paper task second 6 weeks after screen task)	64	Participants took significantly longer to complete the proof-reading task using the computer.
Rice 1994	Computer vs. paper	Comprehension task and text recall measure	Between-Ps	120	Significantly more scoreable unit ideas on paper than in computer mode.
Picking 1997	Computer (three types) vs. static screen representation vs. paper	Proof-reading music scores	Within-Ps (three of five)	19	No significant difference was found between the five presentation styles.
DeAngelis 2000	Computer vs. paper	Examination questions	Within-Ps	30	Participants taking the computer-based examination first scored significantly higher than the paper group.
Hallfors *et al.* 2000	Computer-assisted interview vs. paper	Alcohol, tobacco and drug survey	Between-Ps	2296 (1135 computer; 1161 paper)	Similar outcomes for two conditions, but participants preferred computerised version.
Keogh *et al.* 2000	Computer vs. paper	English language task	Within-Ps	48 (13–14 years)	Children's verbal interactions were mediated by mode of presentation with boys dominating the amount and type of verbal interaction and control of the mouse in the computer task.
Mason *et al.* 2001	Computer vs. paper	Introductory Psychology units	Within-Ps	27	No difference found between the scores for the two conditions.

(*continued*)

Table 2. (*Continued*).

Study	Comparison	Task	Design	Participants	Key Findings
Mayes *et al.* 2001	VDT vs. paper-based reading	Comprehension task with 10 multi-choice items	Between-Ps	40 (Expt. 1)	VDT group took significantly longer to read the article.
Weinberg 2001	Computer vs. paper	Placement test for learning French	Between-Ps	248 (105 computer; 143 paper)	No significant differences found between the two conditions although reactions to the computer version were very positive.
Boyer *et al.* 2002	Internet vs. paper	Survey of Internet purchasing patterns	Between-Ps	416 (155 computer; 261 paper)	Electronic surveys comparable to print surveys, although former had fewer missing responses.
Lee 2002	Composing on computer vs. paper	Timed essays	Within-Ps	6	No conclusion with regard to the two media due to individual differences and small number of participants.
MacCann *et al.* 2002	Computer vs. paper	Free response examination questions	Between-Ps	109 (Expt. 1) (57 computer, 52 paper) 141 (Expt. 2) (88 computer; 53 paper)	No significant differences found between the two media for four of the five questions and authors concluded that not possible to give a clear interpretation for the fifth question.
Knapp and Kirk 2003	Internet vs. touch-tone 'phones vs. paper	Sixty-eight personally sensitive questions	Between-Ps	352	No significant differences for any of the questions across the three media.
Bodmann and Robinson 2004	Computer-based vs. paper tests	Comprehension task with 30 multi-choice items	Between-Ps	55 (Expt 1) (28 computer, 27 paper)	No difference in scores, but completion times significantly longer with paper.

(*continued*)

Table 2. (*Continued*).

Study	Comparison	Task	Design	Participants	Key Findings
Lee 2004	Computer vs. paper-delivered writing test	Essay composition in 50 min	Within-Ps (but all did paper-based test first)	42	No differences in holistic ratings, but analytical components marked higher on computer-generated essays.
Noyes *et al.* 2004	Computers vs. paper	Comprehension task with 10 multi-choice items	Between-Ps	30	No significant difference in comprehension task, but difference in workload with the computer-based test requiring more effort.
Smither *et al.* 2004	Intranet vs. paper-and-pencil	Feedback ratings of line managers	Between-Ps	5257 (48% Intranet; 52% paper)	More favourable ratings from web-based response mode, but controlling for rater and rate characteristics diminished this effect.
Wästlund *et al.* 2005	VDT vs. paper	Comprehension test	Between-Ps	72 (Expt. 1) 72 (Expt. 2)	1. Significantly more correct responses with paper. 2. Significantly greater information production with paper.
Huang 2006	Web-based surveys vs. paper-and-pencil	Security awareness of Internet behaviour	Between-Ps	200	No significant differences found for two media.

VDT = video display terminal.

employing more refined metrics that are more relevant/appropriate for task-specific performance assessment and evaluation.

3.2. *Cognitive indicators*

A consideration of the inherent characteristics of cathode ray tube monitors (flicker, high contrast and fluctuating luminance) would suggest that there should be some differences between performance on screen and that obtained from a comparable paper presentation. Perhaps, therefore, the influence of these features cannot necessarily be measured in the more 'traditional' manner, but requires measures with a greater degree of sensitivity and/or studies that examine performance in the longer term. The physical nature of the computer screen has led to a number of researchers suggesting that lower performance in comparison to paper alternatives might be attributable to additional processing, leading to greater perceived and actual workload (e.g. Ziefle 1998).

In both Noyes and Garland (2003) and Garland and Noyes (2004), differences were found between the two modes of presentation. However, this was in the manner in which the information was retrieved, which suggested that there were differences in memory processing dependent upon the nature of the visual input. Wästlund *et al.* (2005) in a comparison of VDTs and paper found that consumption of information (measured in terms of reading comprehension) and production of information were both superior in the paper-based condition. They provided a psycho-physiological explanation since the VDT condition resulted in a greater level of experienced tiredness and increased feelings of stress. These effects were attributed to an increase in cognitive demands, that is, the computer task was more tiring and more stressful than the paper-based task and this led to 'a greater mobilization of both perceptual and executive cognitive resources' being invested (Wästlund *et al.* 2005, p. 14). Although cognitive workload was not specifically measured, they concluded that the VDT resulted in a higher cognitive workload.

Noyes *et al.* (2004) compared cognitive workload (measured by the NASA-Task Load Index (TLX)) and performance on a comprehension task presented on either computer or paper. Although they found no significant difference in the comprehension scores or the overall workload scores for the two media, significantly more workload was reported on the effort dimension for the computer-based task. This finding is interesting, since it suggests that individuals need to put more effort into the computer task. Hart and Staveland (1988), as part of the development process of the NASA-TLX workload measure, looked at the effect of paper-and-pencil vs. computer administration. Three different tasks were used and it was found that, on average, higher workload was reported when the computer method was used. However, as the correlation between the computer and paper measures was high ($r = 0.94$), Hart and Staveland concluded that this was not a concern when it came to the choice of medium for test administration.

The use of measures such as the NASA-TLX and memory retrieval indicators (Noyes and Garland 2003, Garland and Noyes 2004, Noyes *et al.* 2004) as effective metrics for comparative analysis is supported by other research. For example, Vincent *et al.* (1996) identified relationships between NASA-TLX values and levels of encoding in recognition memory tests and reported higher workload for more deeply encoded information. This suggests that cognitive workload can distinguish differences in processing and that this measure is able to identify small yet important variations in performance, which is especially relevant in more sophisticated tasks that may require sustained attention, decision making, problem solving, etc.

4. Looking to the future

Computer-based versions of 'standardized IQ tests, personality inventories, and behavior and symptom checklist have been in existence for more than 30 years' (Tsemberis *et al.* 1996, p. 167). However, greater availability of computers, increasing use of distance learning through the Internet and wider use of assessment have led to a renewed interest in the computer vs. paper debate with a particular focus on online assessment (see Table 3 for studies relating to online assessment using standardised tests).

Concerns arise from the limited knowledge and understanding of computer-based activities in this arena. As Hargreaves *et al.* (2004) pointed out, young people in the UK today are familiar with using computers and they are also used to undertaking pen and paper assessments. However, a combination of the two activities does not necessarily mean the end result will be satisfactory in providing individuals with a task with which they are comfortable and efficient. When tasks are moved from pen and paper to the computer, equivalence is often assumed, but this is not necessarily the case. For example, even if the paper version has been shown to be valid and reliable, the computer version may not exhibit similar characteristics. If equivalence is required, then this needs to be established.

A large body of literature already exists on online assessment using computers and paper (see, for example, Bodmann and Robinson 2004). Results are mixed; some studies have found benefits relating to computers (e.g. Vansickle and Kapes 1993, Carlbring *et al.* 2007), some have favoured paper (e.g. George *et al.* 1992, Lankford *et al.* 1994, van de Vijver and Harsveld 1994, Russell 1999, McCoy *et al.* 2004), while most have found no differences (e.g. Rosenfeld *et al.* 1992, Kobak *et al.* 1993, Steer *et al.* 1994, King and Miles 1995, DiLalla 1996, Ford *et al.* 1996, Merten and Ruch 1996, Pinsoneault 1996, Smith and Leigh 1997, Neumann and Baydoun 1998, Ogles *et al.* 1998, Donovan *et al.* 2000, Vispoel 2000, Cronk and West 2002, Fouladi *et al.* 2002, Fox and Schwartz 2002, Puhan and Boughton 2004, Williams and McCord 2006). Given this, perhaps decisions concerning the use of computers or paper will need to be made with regard to their various advantages and disadvantages and their relative merits in relation to task demands and required performance outcomes.

4.1. Advantages of online assessment

The benefits of using computers in assessment can be divided into five main categories. These relate to the following.

(1) The richness of the interface. For example, the use of graphics allows a dynamic presentation of the test content. This has the potential to be delivered at various speeds and levels according to the specific needs of the individual. Unlike pen and paper tasks, the computer can make use of the two-way interchange of feedback, that is, not only does the user have feedback from the computer concerning his/her inputs, but the computer is 'learning' about the user from their responses and so can vary the program accordingly. This ability to capture process differences in learners has been cited as one of the major uses of computer-based assessment (Baker and Mayer 1999). Further, computer-based testing also allows other measures relating to cognitive and perceptual performance, for example, reaction time, to be assessed.
(2) The user population. Computer-based testing via the Internet allows a more diverse sample to be located (Carlbring *et al.* 2007) because people only need access

Table 3. Summary of some selected studies concerned with online assessment using standardised tests, from 1992 onwards.

Study	Comparison	Test	Design	Participants	Key Findings
George *et al.* 1992	Computer vs. paper	Beck Depression and Spielberger's State-Anxiety Inventories	Between-Ps	97 (48 computer; 49 paper)	Significant differences found between computer and paper conditions suggesting 'computer anxiety may artificially inflate negative affect scores during computer administration' (p. 203).
Rosenfeld *et al.* 1992	Computer vs. clinician-administered paper version	Yale-Brown Obsessive-Compulsive Scale	Within-Ps	70	High correlations were found between the two modes of administration with participants showing no preference for either version.
Watson *et al.* 1992	Computer vs. paper	Minnesota Multiphasic Personality Inventory	Meta-analyses of nine studies	967 (762 computer administration; 770 paper)	Computer-based scores significantly underestimated booklet counterparts on eight of the 13 scales.
Kobak *et al.* 1993	Computer vs. clinician-administered paper version	Hamilton Anxiety Scale	Within-Ps	292	Computer-administered version demonstrated a strong relationship with the clinician-administered paper version.
Mead and Drasgow 1993	Computerised vs. paper-and-pencil tests	Timed power and speeded tests of cognitive abilities	Meta-analysis of 28 studies		Strong support for there being no effect of medium for carefully constructed power tests, and a substantial effect for speeded tests.
Vansickle and Kapes 1993	Computer vs. paper	Strong-Campbell Interest Inventory	Between-Ps	52	Computer version was found to be more reliable and faster.
Finegan and Allen 1994	Computerised and written questionnaires	Personality measures	Between-Ps	63 (Expt. 1) (31 computer; 32 written)	Participants in the computer condition expressed significantly higher achievement motivation and socially desirable responses.

(*continued*)

Table 3. (*Continued*).

Study	Comparison	Test	Design	Participants	Key Findings
			Within-Ps	40 (Expt. 2)	Participants showed stronger opposition towards censorship and more positive work attitudes on the computerised version.
Lankford *et al.* 1994	Computer vs. paper	Beck Depression Inventory, Purpose In Life, and three anxiety questionnaires	Between-Ps	131 (65 computer; 66 paper)	Results suggested use of computer-administered personality tests may not be valid.
Steer *et al.* 1994	Computer vs. paper	Beck Depression Inventory, and Hopelessness Scale	Between-Ps (previous paper results)	330 (all computer)	Computer-administered test results were comparable with those previously reported for the printed versions.
Van de Vijver and Harsveld 1994	Computerised version of GATB vs. paper-and-pencil	General Aptitude Test Battery (GATB)	Between-Ps	326	Simple tests more affected by computerisation with significantly faster and more inaccurate responses.
King and Miles 1995	Computer vs. paper	Four work-related measures	Within-Ps	674	No difference found between computerised and paper versions.
Locke and Gilbert 1995	Computer vs. questionnaire vs. interview	Minnesota Multiphasic Personality Inventory and Drinking Habits Questionnaire	Between-Ps	162	Lower levels of self-disclosure occurred in the interview condition.
DiLalla 1996	Computer vs. paper	Multidimensional Personality Questionnaire	Between-Ps	227 (126 computer; 101 paper)	No significant differences found between the two media.
Ford *et al.* 1996	Computer vs. paper	Four personality scales	Between-Ps	52	No significant differences found between the two media.

(*continued*)

Table 3. (*Continued*).

Study	Comparison	Test	Design	Participants	Key Findings
Merten and Ruch 1996	Computer vs. paper	Eysenck Personality Questionnaire and Carroll Rating Scale for Depression	Within-Ps	72 (four groups: mode and test order)	No systematic difference found.
Peterson *et al.* 1996	Computer vs. paper	Beck Depression Inventory, and mood and intelligence tests	Between-Ps	57 (condition numbers not given)	Mean test results comparable for the two versions; however, a tendency for the computer test to yield higher scores on the Beck Depression Inventory.
Pinsoneault 1996	Computer vs. paper	Minnesota Multiphasic Personality Inventory-2	Within-Ps	32	Two formats were comparable, but participants expressed preference for the computer version.
Webster and Compeau 1996	Computer vs. paper	Computer experience and playfulness questionnaires	Between-Ps	95 (45 computer; 50 paper)	No difference found in administration mode, but differences in nomological validity found.
Hansen *et al.* 1997	Computer vs. paper	Strong Interest Inventory	Between-Ps	155 (71 computer; 84 paper)	Slight preference for computers, and paper-and-pencil group had significantly more difficulty using the test materials.
Potosky and Bobko 1997	Computer vs. paper	Personal Experiences and Attitudes Questionnaire	Within-Ps	176	No difference found between computerised and paper versions, and mixed findings concerning response distortion measures.
Smith and Leigh 1997	Internet vs. paper	Human sexuality measures	Between-Ps	128 (72 computer; 56 paper)	Response patterns similar for the two groups.

(*continued*)

Table 3. (*Continued*).

Study	Comparison	Test	Design	Participants	Key Findings
Tseng *et al*. 1997	Computer vs. paper	Computer anxiety and mood assessment questionnaires	Between-Ps	157 (54 computer; 54 paper; 49 control)	Self-ratings of anxiety and mood were found to be significantly greater in the computer than the paper-and-pencil groups.
Miles and King 1998	Computer vs. paper	Four personality tests	Between-Ps	874 (483 computer; 391 paper)	Significant differences for administration mode and gender were found, but no interactions.
Neuman and Baydoun 1998	Computer vs. paper	Office Skills Tests	Within-Ps	411	No differences found between the two conditions.
Ogles *et al*. 1998	Computer vs. paper	Four depression inventories	Within-Ps	113	Psychometric characteristics of computer version maintained the standards of the paper and pencil version.
Schwartz *et al*. 1998	Computer vs. paper	Ego Identity Process questionnaire and Identity Style Inventory	Between-Ps	213 (100 computer; 113 paper)	Significant differences in identity status found between two conditions.
Tseng *et al*. 1998	Computer vs. Personal Digital Assistant (PDA) or paper	Visual Analogue Mood Scales	Between-Ps	136 (47 computer; 43 PDA; 46 paper)	No difference in mood scores across the three conditions, but found computer anxiety can affect the results, and suggested PDAs may have advantages over computers.
Davis 1999	Web vs. paper	Ruminative Responses personality questionnaire	Between Ps	1371 (128 web; 118 + 113 + 1012 paper)	Results suggested web was comparable with paper test administration.
Donovan *et al*. 2000	Computer vs. paper	Three standardised job scales	Between-Ps	2286 (509 computer; 1777 paper)	Results supported measurement equivalence.

(*continued*)

Table 3. (*Continued*).

Study	Comparison	Test	Design	Participants	Key Findings
Vispoel 2000	Computer vs. paper	Self-Description Questionnaire	Within-Ps	212	Results supported the comparability of scores across the two media.
Vispoel *et al.* 2001	Computer vs. paper	Rosenberg Self-Esteem Scale	Within-Ps	224	Mode had little effect on psychometric properties (i.e. score magnitude, variability, and factor structure), but computerised version took longer and was preferred.
Cronk and West 2002	Internet vs. paper either take-home or in-class	Visions of Morality 30-item scale	Between-Ps	236 (59 per condition)	No significant differences in scores between four conditions.
Fouladi *et al.* 2002	Internet vs. paper	Three standardised questionnaires on mood	Between-Ps	398 (234 computer; 164 paper)	Magnitude of mode effects was very small; authors concluded that online administration was equivalent.
Fox and Schwartz 2002	Computer vs. paper	Personality questionnaires	Between-Ps	200	No significant differences found between the two media.
Pomplun *et al.* 2002	Computer vs. paper	Nelson-Denny Reading Test	Between-Ps	215 (94 computer; 121 paper)	Small or no differences found between the two media.
McCoy *et al.* 2004	Electronic vs. paper surveys	Technology Acceptance Model (TAM) instrument	Within-Ps (but all did electronic form first)	120	Significantly lower responses were made on the electronic forms.
Puhan and Boughton 2004	Computerised version of teacher certification test vs. paper and pencil	Teacher certification test: reading, writing and mathematics	Between-Ps (free to choose medium)	7308 (3308 computer; 4000 paper)	Two versions of test were comparable.

(*continued*)

Table 3. (*Continued*).

Study	Comparison	Test	Design	Participants	Key Findings
Wolfe and Manalo 2004	Word processing vs. handwriting	Essay composition (as part of English as a Second Language [ESL] test)	Between-Ps	133,906 (54% word processor; 46% handwriting)	Participants with lower levels of proficiency in English receive higher marks for handwritten essays.
Booth-Kewley *et al.* 2005	Computer vs. paper	Three scales on sensitive behaviours	Between-Ps	301 (147 computer; 154 paper)	Computer administration produced a sense of disinhibition in respondents.
Carlbring *et al.* 2005	Internet vs. paper	Eight clinical (panic disorder) questionnaires	Between-Ps	494	Significantly higher scores found for Internet versions of three of eight questionnaires.
Williams and McCord 2006	Computer vs. paper	Raven Standard Progressive Matrices	Mixed	50 (12 computer-computer; 11 paper-paper; 12 computer-paper; 15 paper-computer)	Tentative support for the equivalence of the two versions.

to a computer. It also lets people take part in testing from their homes; this group may not necessarily be available for testing in a laboratory setting due to mobility and, perhaps, disability issues.

(3) Standardisation of test environment, that is, the test is presented in the same way and in the same format for a specified time. Thus, errors in administration, which may lead to bias, are minimised (Zandvliet and Farragher 1997). Another benefit of computer responding is that the subjective element attached to handwritten responses is removed.

(4) Online scoring. This results in faster feedback and greater accuracy, that is, reduction in human error. Information relating to test-taking behaviours, for example, how much time was spent on each item, can be readily collected (Liu *et al.* 2001). It is generally accepted that delivery and scoring of the test online leads to economic cost savings, especially for large samples.

(5) Quantity and quality of composition. Goldberg *et al.* (2003) in a meta-analysis of studies comparing computers and paper over a 10-year period found significant effect sizes, which favoured computers in both quantity and quality of writing. It should be noted that there is a large literature on differences in composition, and quality and assessment, of written output on computer and paper (Lee 2002, MacCann *et al.* 2002), which is not being covered here.

4.2. Disadvantages of using computers in assessment

One of the primary concerns of using computers is the need for computer proficiency and typing skills (Wang and Kolen 2001, Gallagher *et al.* 2002). As a result, some participants may not feel comfortable with the computer medium; Wolfe and Manalo (2004) in their study addressed this by having participants select their response mode. Further, they may need more time as demonstrated by Zandvliet and Farragher (1997), who found that individuals with minimal computer skills took longer. More specific disadvantages relate to:

(1) Lack of a controlled environment with responses being made at various times and settings and perhaps not even by the designated individual (Fouladi *et al.* 2002). Double submissions may also be a problem as found by Cronk and West (2002).

(2) Computer hardware and software. These may be subject to freezing and crashing; in the test setting, time can be wasted when computers have to be restarted or changed. As an example, Zandvliet and Farragher (1997) found that a computer-administered version of a test required more time than the written test. Further, there may be differences in the layout of the test depending on a respondent's particular browser software and settings.

(3) The computer screen. For longer tests, it may be more tiring to work on the computer than on paper (Ziefle 1998).

(4) Serial presentation. It is difficult to attain equivalence with computer and paper presentation. As a general rule, it is easier to look through items and move backwards and forwards when using paper.

(5) Concerns about confidentiality. This is particularly the case with Internet-based assessments (Morrel-Samuels 2003) and raises a number of ethical and clinical issues. For example, there is a tendency for respondents to create a particular impression in their results. This social desirability effect has been found to be more

likely in web surveys (Morrel-Samuels 2003) and has been the subject of much interest in the administration of personality questionnaires (see Ford *et al*. 1996, Fox and Schwartz 2002). Participants are often more disinhibited in computer-based tests and will report more risky behaviours (Locke and Gilbert 1995, Booth-Kewley *et al*. 2007). The latter is not necessarily a disadvantage; the issue relates to the degree to which the respondents are being honest and a valid response is being attained.

Some of these disadvantages can be addressed, for example, through availability of a reversion capability and password access and unique codes to mask the identity of respondents. Carlbring *et al*. (2007) put forward a number of criteria for computer-based questions; they stated that questions need to be easily understood and brief and only require single responses. The issue of social desirability is not so easy to address. This is primarily because, in the large body of literature on this topic, there is little consensus (see King and Miles 1995, Booth-Kewley *et al*. 2007).

4.3. User preferences for the two media

When comparing studies across Tables 1–3, it becomes apparent that user preferences are being noted in the more recent studies, especially those relating to standardised test administration in Table 3. For example, Pinsoneault (1996), Hansen *et al*. (1997), Vispoel (2000), Vispoel *et al*. (2001) all indicated that participants preferred the computer form of the test. Likewise, Horton and Lovitt (1994) and Hallfors *et al*. (2000) found that participants favoured learning information from a computer and were more positive towards computers than paper. Dillon (1992, p. 1305) concluded that 'hardcopy seems to be preferred to screen displays', although he noted Muter and Mauretto's (1991) finding that preferences are shifting as screen technology improves. Perhaps the point has now been reached where computer technology has improved sufficiently to allow sophisticated manipulation facilities for ensuring greater computer and paper equivalence and this is being reflected in people's preferences.

Online assessment has traditionally used standardised tests. One of the early concerns related to the need to be able to use a computer and some studies (George *et al*. 1992, Tseng *et al*. 1997) have noted the increased computer anxiety associated with online assessment. However, as the population becomes more computer literate, this concern should fade. In fact, Hansen *et al*. (1997) found that the paper-and-pencil group had more difficulty using the test materials. User preference certainly seems to support the use of online assessment.

5. General discussion

In his general conclusion, Dillon (1992, pp. 1322–1323) called for a move away from the 'rather limited and often distorted view of reading', to 'a more realistic conceptualisation'. When considering 'paper and screens', this appears to have happened as more traditional indicators such as reading speed have been replaced with more sophisticated measures such as cognitive workload and memory measures.

The post-1992 literature review distinguishes three types of task in the computer and paper comparisons. These are non-standardised, open-ended tasks (for example, composition), non-standardised, closed tasks (for example, multi-choice questions) and standardised tasks (for example, the administration of standardised tests). In terms of

equivalence, the issue relates to whether a task in paper form remains the same when transferred to a computer. However, equivalence has been defined in different ways in the research literature according to the type and extent of psychometric analyses being applied (Schulenberg and Yutrzenik 1999). This may help explain why findings are mixed with some studies indicating equivalence and some not. This applies to both traditional indicators and more sophisticated measures.

On a pragmatic level, equivalence is going to be hard to achieve since two different presentation and response modes are being used. This will especially be the case with non-standardised, open-ended tasks. As Lee (2004) found, participants spent significantly longer in the pre-writing planning stage when using paper than when using a computer. In contrast, Dillon's (1992) paper focused on the second type of task. It would seem that when the task is bespoke and closed it can be designed to ensure that computer and paper presentations are made as similar as possible. Thus, experiential equivalence can be more readily achieved today, whereas prior to 1992 computers lacked this flexibility. Finally, the third group comprising standardised test administration is probably the most likely to be lacking in equivalence. For example, Finegan and Allen (1994), Lankford *et al.* (1994), Peterson *et al.* (1996), Tseng *et al.* (1997), Schwartz *et al.* (1998) and Carlbring *et al.* (2007) all found that the computerised versions of their tests produced significantly different results from their paper counterparts. This is because this type of test has often been devised and administered on paper and then transferred in the same form to a computer in order to ensure that respondents are completing a direct replication of the task. However, the psychometric properties (norms, distributions, reliability, validity) of the standardised tests in paper form are often well-established and can therefore be checked against the computerised version (see Williams and McCord 2006). Thus, psychometric equivalence can be readily gauged, if not achieved.

A further development from the situation in 1992 is the use of the Internet for test administration and, as seen from Table 3, there has been much research interest in comparing web- and paper-based testing. However, reference to the Internet as a mode of administration is really a distraction because, with a known respondent population as is likely with standardised tests, the issue relates to using a computer instead of hard copy. Hence, the same issues relating to hardware and software platforms, and achieving experiential and psychometric equivalence, exist.

In conclusion, the achievement of equivalence in computer- and paper-based tasks poses a difficult problem and there will always be some tasks where this is not possible. However, what is apparent from this review is that the situation is changing and it is probably fair to conclude that greater equivalence is being achieved today than at the time of Dillon's (1992) literature review.

References

Askwall, S., 1985. Computer supported reading vs reading text on paper: A comparison of two reading situations. *International Journal of Man-Machine Studies*, 22, 425–439.

Baker, E.L. and Mayer, R.E., 1999. Computer-based assessment of problem solving. *Computers in Human Behavior*, 15, 269–282.

Belmore, S.M., 1985. Reading computer-presented text. *Bulletin of the Psychonomic Society*, 23, 12–14.

Bodmann, S.M. and Robinson, D.H., 2004. Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research*, 31, 51–60.

Booth-Kewley, S., Larson, G.E., and Miyoshi, D.K., 2007. Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior*, 23, 463–477.

Boyer, K.K., *et al*., 2002. Print versus electronic surveys: A comparison of two data collection methodologies. *Journal of Operations Management*, 20, 357–373.
Carlbring, P., *et al*., 2007. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, 23, 1421–1434.
Creed, A., Dennis, I., and Newstead, S., 1987. Proof-reading on VDUs. *Behaviour & Information Technology*, 6, 3–13.
Cronk, B.C. and West, J.L., 2002. Personality research on the Internet: A comparison of Web-based and traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments and Computers*, 34 (2), 177–180.
Cushman, W.H., 1986. Reading from microfiche, VDT and the printed page. *Human Factors*, 28, 63–73.
Davis, R.N., 1999. Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments and Computers*, 31 (4), 572–577.
DeAngelis, S., 2000. Equivalency of computer-based and paper-and-pencil testing. *Journal of Allied Health*, 29, 161–164.
DiLalla, D.L., 1996. Computerized administration of the Multidimensional Personality Questionnaire. *Assessment*, 3, 365–374.
Dillon, A., 1992. Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, 35, 1297–1326.
Dillon, A., 1994. *Designing usable electronic text: Ergonomic aspects of human information usage*. London: Taylor & Francis.
Donovan, M.A., Drasgow, F., and Probst, T.M., 2000. Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. *Journal of Applied Psychology*, 85, 305–313.
Finegan, J.E. and Allen, N.J., 1994. Computerized and written questionnaires: Are they equivalent? *Computers in Human Behavior*, 10, 483–496.
Ford, B.D., Vitelli, R., and Stuckless, N., 1996. The effects of computer versus paper-and-pencil administration on measures of anger and revenge with an inmate population. *Computers in Human Behavior*, 12, 159–166.
Fouladi, R.T., McCarthy, C.J., and Moller, N.P., 2002. Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9, 204–215.
Fox, S. and Schwartz, D., 2002. Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computers in Human Behavior*, 18, 389–410.
Gallagher, A., Bridgeman, B., and Cahalan, C., 2002. The effect of computer-based tests on racial-ethnic and gender groups. *Journal of Educational Measurement*, 39, 133–147.
Garland, K.J. and Noyes, J.M., 2004. CRT monitors: Do they interfere with learning? *Behaviour & Information Technology*, 23, 43–52.
George, C.E., Lankford, J.S., and Wilson, S.E., 1992. The effects of computerized versus paper-and-pencil administration on negative affect. *Computers in Human Behavior*, 8, 203–209.
Goldberg, A., Russell, M., and Cook, A., 2003. The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, 2, 1–51. [online]. Available from: http://www.jtla.org [Accessed 15 June 2006].
Gould, J.D., 1981. Composing letters with computer-based text editors. *Human Factors*, 23, 593–606.
Gould, J.D. and Grischkowsky, N., 1984. Doing the same work with hard copy and with Cathode-Ray Tube (CRT) computer terminals. *Human Factors*, 26, 323–337.
Gould, J.D., *et al*., 1987a. Reading is slower from CRT displays than the paper: Attempts to isolate a single-variable explanation. *Human Factors*, 29, 269–299.
Gould, J.D, *et al*., 1987b. Reading from CRT displays can be as fast as reading from paper. *Human Factors*, 29, 497–517.
Gray, S.H., Barber, C.B., and Shasha, D., 1991. Information search with dynamic text vs paper text: An empirical comparison. *International Journal of Man-Machine Studies*, 35, 575–586.
Hallfors, D., *et al*., 2000. A comparison of paper vs. computer-assisted self interview for school alcohol, tobacco, and other drug surveys. *Evaluation and Program Planning*, 23, 149–155.
Hansen, J.-I.C., *et al*., 1997. Comparison of user reaction to two methods of Strong Interest Inventory administration and report feedback. *Measure and Evaluation in Counseling and Development*, 30, 115–127.

Hansen, W.J., Doring, R., and Whitlock, L.R., 1978. Why an examination was slower on-line than on paper? *International Journal of Man-Machine Studies*, 10, 507–519.
Hargreaves, M., *et al*., 2004. Computer or paper? That is the question: Does the medium in which assessment questions are presented affect children's performance in mathematics? *Educational Review*, 46, 29–42.
Hart, S.G. and Staveland, L.E., 1988. Development of the NASA-TLX (Task Load Index): Results of experimental and theoretical research. *In*: P.A. Hancock and N. Meshkati, eds. *Human mental workload*. Amsterdam: North-Holland, 139–183.
Heppner, F.H., *et al*., 1985. Reading performance on a standardized text is better from print than computer display. *Journal of Reading*, 28, 321–325.
Horton, S.V. and Lovitt, T.C., 1994. A comparison of two methods of administrating group reading inventories to diverse learners. *Remedial and Special Education*, 15, 378–390.
Huang, H.-M., 2006. Do print and Web surveys provide the same results? *Computers in Human Behavior*, 22, 334–350.
Kak, A., 1981. Relationship between readability of printed and CRT-displayed text. *In*: *Proceedings of the 25th annual meeting of the Human Factors Society*. Santa Monica, CA: Human Factors and Ergonomics Society, 137–140.
Keenan, S.A., 1984. Effects of chunking and line length on reading efficiency. *Visible Language*, 18, 61–80.
Keogh, T., *et al*., 2000. Gender, pair composition and computer versus paper presentations of an English language task. *Educational Psychology*, 20, 33–43.
King, W.C. and Miles, E.W., 1995. A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology*, 80, 643–651.
Knapp, H. and Kirk, S.A., 2003. Using pencil and paper, Internet and touch-tone phones for self-administered surveys: Does methodology matter? *Computers in Human Behavior*, 19, 117–134.
Kobak, K.A., Reynolds, W.M., and Greist, J.H., 1993. Development and validation of a computer-administered version of the Hamilton Anxiety Scale. *Psychological Assessment*, 5, 487–492.
Kruk, R.S. and Muter, P., 1984. Reading continuous text on video screens. *Human Factors*, 26, 339–345.
Lankford, J.S., Bell, R.W., and Elias, J.W., 1994. Computerized versus standard personality measures: Equivalency, computer anxiety, and gender differences. *Computers in Human Behavior*, 10, 497–510.
Lee, H.K., 2004. A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing*, 9, 4–26.
Lee, Y.-J., 2002. A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessing Writing*, 8, 135–157.
Liu, M., Papathanasiou, E., and Hao, Y.-W., 2001. Exploring the use of multimedia examination formats in undergraduate teaching: Results from the fielding testing. *Computers in Human Behavior*, 17, 225–248.
Locke, S.D. and Gilbert, B.O., 1995. Method of psychological assessment, self-disclosure, and experiential differences: A study of computer, questionnaire, and interview assessment formats. *Journal of Social Behavior and Personality*, 10, 255–263.
Lukin, M.E., *et al*., 1985. Comparing computerized versus traditional psychological assessment. *Computers in Human Behavior*, 1, 49–58.
MacCann, R., Eastment, B., and Pickering, S., 2002. Responding to free response examination questions: Computer versus pen and paper. *British Journal of Educational Technology*, 33, 173–188.
McCoy, S., *et al*., 2004. . Electronic versus paper surveys: Analysis of potential psychometric biases. *In*: *Proceedings of the IEEE 37th Hawaii International Conference on System Science*. Washington, DC: IEEE Computer Society, 80265C.
Mason, B.J., Patry, M., and Bernstein, D.J., 2001. An examination of the equivalence between non-adaptive computer-based and traditional testing. *Journal of Educational Computing Research*, 24, 29–39.
Mayes, D.K., Sims, V.K., and Koonce, J.M., 2001. Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, 28, 367–378.
Mead, A.D. and Drasgow, F., 1993. Equivalence of computerized and paper and pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449–458.

Merten, T. and Ruch, W., 1996. A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for Depression. *Personality and Individual Differences*, 20, 281–291.

Miles, E.W. and King, W.C., 1998. Gender and administration mode effects when pencil-and-paper personality tests are computerized. *Educational and Psychological Measurement*, 58, 68–76.

Morrel-Samuels, P., 2003. Web surveys' hidden hazards. *Harvard Business Review*, 81, 16–18.

Muter, P., *et al*., 1982. Extended reading of continuous text on television screens. *Human Factors*, 24, 501–508.

Muter, P. and Maurutto, P., 1991. Reading and skimming from computer screens and books: The paperless office revisited?. *Behaviour & Information Technology*, 10, 257–266.

Neumann, G. and Baydoun, R., 1998. Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement*, 22, 71–83.

Noyes, J.M. and Garland, K.J., 2003. VDT versus paper-based text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics*, 31, 411–423.

Noyes, J.M., Garland, K.J., and Robbins, E.L., 2004. Paper-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology*, 35, 111–113.

Oborne, D.J. and Holton, D., 1988. Reading from screen versus paper: There is no difference. *International Journal of Man-Machine Studies*, 28, 1–9.

Ogles, B.M., *et al*., 1998. Computerized depression screening and awareness. *Community Mental Health Journal*, 34, 27–38.

Oliver, R., 1994. Proof-reading on paper and screens: The influence of practice and experience on performance. *Journal of Computer-Based Instruction*, 20, 118–124.

Peterson, L., Johannsson, V., and Carlsson, S.G., 1996. Computerized testing in a hospital setting: Psychometric and psychological effects. *Computers in Human Behavior*, 12, 339–350.

Picking, R., 1997. Reading music from screens vs paper. *Behaviour & Information Technology*, 16, 72–78.

Pinsoneault, T.B., 1996. Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Inventory-2. *Computers in Human Behavior*, 12, 291–300.

Pomplun, M., Frey, S., and Becker, D.F., 2002. The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62, 337–354.

Potosky, D. and Bobko, P., 1997. Computer versus paper-and-pencil administration mode and response distortion in noncognitive selection tests. *Journal of Applied Psychology*, 82, 293–299.

Puhan, G. and Boughton, K.A., 2004. Evaluating the comparability of paper and pencil versus computerized versions of a large-scale certification test. *In*: *Proceedings of the annual meeting of the American Educational Research Association (AERA)*. San Diego, CA: Educational Testing Service.

Rice, G.E., 1994. Examining constructs in reading comprehension using two presentation modes: Paper vs. computer. *Journal of Educational Computing Research*, 11, 153–178.

Rosenfeld, R., *et al*., 1992. A computer-administered version of the Yale-Brown Obsessive-Compulsive Scale. *Psychological Assessment*, 4, 329–332.

Russell, M., 1999. Testing on computers: A follow-up study comparing performance on computer and paper. *Education Policy Analysis Archives*, 7, 20.

Schulenberg, S.E. and Yutrzenk, B.A., 1999. The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. *Behavior Research Methods, Instruments, & Computers*, 31, 315–321.

Schwartz, S.J., Mullis, R.L., and Dunham, R.M., 1998. Effects of authoritative structure in the measurement of identity formation: Individual computer-managed versus group paper-and-pencil testing. *Computers in Human Behavior*, 14, 239–248.

Sellen, A.J. and Harper, R.H.R., 2002. *The myth of the paperless office*. Cambridge, MA: The MIT Press.

Smith, M.A. and Leigh, B., 1997. Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments and Computers*, 29, 496–505.

Smither, J.W., Walker, A.G., and Yap, M.K.T., 2004. An examination of the equivalence of Web-based versus paper-and-pencil upward feedback ratings: Rater- and rate-level analysis. *Educational and Psychological Measurement*, 64, 40–61.

Steer, R.A., *et al*., 1994. Use of the computer-administered Beck Depression Inventory and Hopelessness Scale with psychiatric patients. *Computers in Human Behavior*, 10, 223–229.
Switchenko, D., 1984. Reading from CRT versus paper: The CRT-disadvantage hypothesis re-examined. *In*: *Proceedings of the 28th annual meeting of the Human Factors Society*. Santa Monica, CA: Human Factors and Ergonomics Society, 429–430.
Tsemberis, S., Miller, A.C., and Gartner, D., 1996. Expert judgments of computer-based and clinician-written reports. *Computers in Human Behavior*, 12, 167–175.
Tseng, H.-M., Macleod, H.A., and Wright, P.W., 1997. Computer anxiety and measurement of mood change. *Computers in Human Behavior*, 13, 305–316.
Tseng, H.-M., *et al*., 1998. Computer anxiety: A comparison of pen-based digital assistants, conventional computer and paper assessment of mood and performance. *British Journal of Psychology*, 89, 599–610.
van De Velde, C. and von Grünau, M., 2003. Tracking eye movements while reading: Printing press versus the cathode ray tube. *In*: *Proceeding of the 26th European conference on visual perception, ECVP 2003*. Paris, France: ECVP, 107.
van De Vijver, F.J.R. and Harsveld, M., 1994. The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology*, 79, 852–859.
Vansickle, T.R. and Kapes, J.T., 1993. Comparing paper-pencil and computer-based versions of the Strong-Campbell Interest Inventory. *Computers in Human Behavior*, 9, 441–449.
Vincent, A., Craik, F.I.M., and Furedy, J.J., 1996. Relations among memory performance, mental workload and cardiovascular responses. *International Journal of Psychophysiology*, 23, 181–198.
Vispoel, W.P., 2000. Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preferences. *Measurement and Evaluation in Counseling and Development*, 33, 130–143.
Vispoel, W.P., Boo, J., and Bleiler, T., 2001. Computerized and paper-and-pencil versions of the Rosenberg Self-Esteem Scale: A comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement*, 61, 461–474.
Wang, T. and Kolen, M.J., 2001. Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. *Journal of Educational Measurement*, 38, 19–49.
Wästlund, E., *et al*., 2005. Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior*, 21, 377–394.
Watson, C.G., Thomas, D., and Anderson, P.E.D., 1992. Do computer-administered Minnesota Multiphasic Personality Inventories underestimate booklet-based scores? *Journal of Clinical Psychology*, 48, 744–748.
Webster, J. and Compeau, D., 1996. Computer-assisted versus paper-and-pencil administration of questionnaires. *Behavior Research Methods, Instruments and Computers*, 28, 567–576.
Weinberg, A., 2001. Comparison of two versions of a placement test: Paper-pencil version and computer-based version. *Canadian Modern Language Review*, 57, 607–627.
Weldon, L.J., *et al*., 1985. . The structure of information in online and paper technical manuals. *In*: R.W. Swezey, ed. *Proceedings of the 29th annual meeting of the Human Factors Society*, Vol. 2. Santa Monica, CA: Human Factors Society, 1110–1113.
Wilkinson, R.T. and Robinshaw, H.M., 1987. Proof-reading: VDU and paper text compared for speed, accuracy, and fatigue. *Behaviour & Information Technology*, 6, 125–133.
Williams, J.E. and McCord, D.M., 2006. Equivalence of standard and computerized versions of the Raven Progressive Matrices test. *Computers in Human Behavior*, 22, 791–800.
Wolfe, E.W. and Manalo, J.R., 2004. Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning and Technology*, 8, 53–65.
Wright, P. and Lickorish, A., 1983. Proof-reading texts on screen and paper. *Behaviour and Information Technology*, 2, 227–235.
Zandvliet, D. and Farragher, P., 1997. A comparison of computer-administered and written tests. *Journal of Research on Computing in Education*, 29, 423–438.
Ziefle, M., 1998. Effects of display resolution on visual performance. *Human Factors*, 40, 554–568.