THE PSYCHOMETRIC EVALUATION OF EDUCATIONAL INTRANETS

PAUL VAN SCHAIK
JONATHAN LING
University of Teesside

ABSTRACT

A major determinant of the success of educational intranet sites is their usability. In addition to measures of task performance and navigation behavior, psychometric instruments are used to evaluate usability and the quality of human-computer interaction more generally. However, there is a lack of validated instruments for the evaluation of sites. The aim of the present study was to establish the psychometric properties of scales for measuring the quality of interaction with educational intranets. In two experiments, undergraduates ($n = 81$ and $n = 109$, respectively) performed an information retrieval task using different versions of the same university intranet site. Psychometric properties of the instruments were analyzed. The quality of the scales is evaluated, recommendations for their use are given, and further research into the scales is outlined.

INTRODUCTION

The use of intranets is growing in numerous domains, including business (Murgolo-Poore, Pitt, & Ewing, 2002), medicine (Stalidis, Prentza, Vlachos, Maglavera, & Koutsouris, 2001), and higher education (Browning & Williams, 1997). Intranets are computing networks that are used internally by an organization to disseminate important or useful information to its members; typically an intranet is implemented as an internal Web, based on the same technology as external World Wide Web (Web) sites. For example, universities often employ intranets that include various types of learning material as a resource for students.

81

Bespoke and commercially available Virtual Learning Environments (VLEs), such as Blackboard™ and WebCT™, represent a dedicated type of intranet and are used increasingly to support learning and teaching in higher education. Unfortunately, one common problem for intranets is that of poor design (Nielsen, 1999); therefore, there is a need for organizations to improve the usability of intranets in order to boost productivity (Nielsen, 2001).

Good measurement instruments are essential and a prerequisite for establishing if design changes result in any improvements in intranet usability. Three main types of measure are frequently used to assess the quality of human-computer interaction: psychometric, task performance, and behavioral measures. Psychometric instruments are commonly used to measure the quality of interaction as perceived by users (Lewis, 2002). Although this type of instrument is increasingly used for online data collection (van Schaik & Ling, 2003a), its use is only appropriate if an instrument possesses good psychometric properties. In particular, factor structure, reliability, validity, and sensitivity are commonly used to establish the quality of psychometric instruments (Lewis, 2002). Measures of task performance (e.g., accuracy and speed) and behavioral measures (e.g., users' navigation through a Web site in terms of number of page visits) are also typically used. The current study used both of the latter to validate psychometric measures.

Psychometric properties of the following usability questionnaires and their use in usability evaluation have ben published widely: QUIS (Chin, Diehl, & Norman, 1988), PSSUQ (Lewis, 2002), and SUMI (Kirakowski, 1996). These instruments have been developed for use with human-computer interfaces in general, but not specifically for the evaluation of Web sites, and all have good psychometric properties (Chin et al., 1988; Lewis, 2002; Kirakowski, 1996). Although the SUS (Brooke, 1996) is also widely used to evaluate user interfaces, less information is available on its properties. Constructs measured by these instruments include system usefulness, information quality, interface quality (PSSUQ), screen, terminology and system information, learning, system capabilities (QUIS), efficiency, affect, helpfulness, control, and learnability (SUMI). In addition, all four instruments include an overall measure of usability. The WAMMI questionnaire has been developed for measuring users' satisfaction with Web sites (Kirakowski & Kierik, 1998). The questionnaire is divided into five reliable subscales that are very similar to those of the SUMI (Kirakowski, 1996).

Unfortunately, there is a lack of validated theoretically-based key measures of interaction with educational intranets. However, a number of measures that have been used for other purposes is available and could be used, if validated in the context of educational intranets. The concept of disorientation has been specifically linked with navigation of hypertext and hypermedia systems in general (Conklin, 1987) and Web sites in particular (Ahuja & Webster, 2001). Another key concept, aesthetic value, has been linked to the usability of user interfaces (Tractinsky, Katz, & Ikar, 2000) and Web sites in particular (Lindagaard & Dudek, 2003). The concept of flow—a psychological condition in

which a person feels cognitively efficient, motivated, and happy (Moneta & Csikszentmihalyi, 1996)—has been identified as a further key concept in the quality of human-computer interaction (Chen, Wigand, & Nilan, 1999; Davis & Wiedenbeck, 2001). Technology acceptance is arguably one of the most important considerations in the quality of interaction because potential improvements in task performance will be lost if users fail to accept the very interactive computer systems that could in principle bring about improved usability (Davis, 1993; Thong, Hong, & Tam, 2002). An advantage of theory-based measures (such as perceived usefulness within the framework of the Technology Acceptance Model; see Davis & Wiedenbeck, 2001) is that they can be used to further test and refine models of human-computer interaction and derive theory-based, and therefore generic, implications for system re-design.

Currently, there is a lack of validated psychometric measures for evaluating Web sites, and this shortcoming is particularly evident for educational intranets. In two experiments, the current study investigates the psychometric properties of a set of existing instruments in the context of an educational intranet. A realistic task scenario was studied, using an experimental paradigm mimicking HE students' interaction in actual intranet use.

EXPERIMENT 1

Method

Experimental Design and Materials

An independent measures design with two experimental conditions was used in order to test the sensitivity of psychometric measures, as well as other psychometric properties. Link color was the independent variable. The two levels were blue (the default link color in Web browsers), and a combination of black for global (hyper)links (having the highest contrast with the default background color of white) and maroon for in-text links (to make these distinguishable from text presented in the default text color of black, but still having good contrast with a white background.

An existing intranet site, developed by the University of Teesside Psychology department for its students, was used. The site consisted of a home page (see Figure 1), six main content pages, a sitemap, a search page, and several hundred further content pages. This site included information such as learning resources from taught modules, generic study information, careers advice, and staff details as well as links to other university intranet sites and external Internet sites. Previous research (Campbell & Maglio, 1999) found that blue hyperlinks produced superior performance and subjective outcomes when underlined and in bold. Therefore, of the two versions of the site that were produced, one had blue hyperlinks in bold (HTML code: #0000FF) both for global navigation and other navigation; the other had black hyperlinks (HTML code: #000000) for global
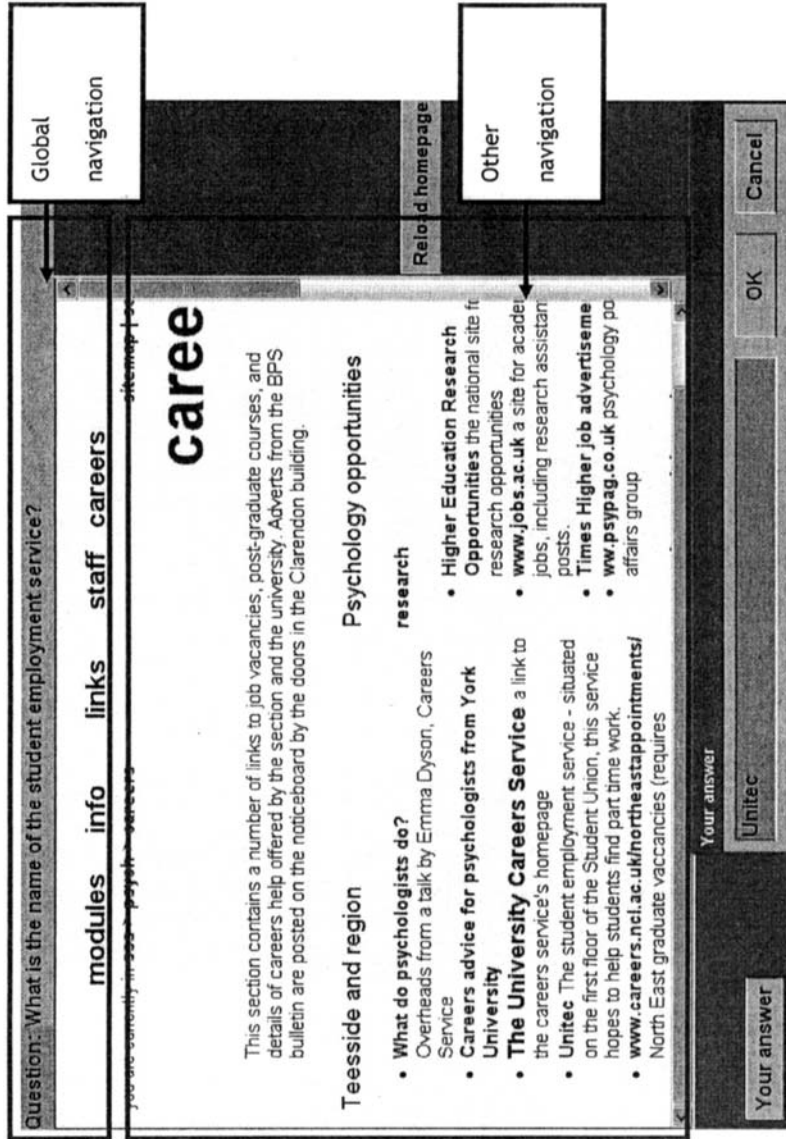
Figure 1. A typical Web page used in Experiment 1 (link color: black and maroon).

navigation and maroon hyperlinks (HTML code: #800000) for other navigation. Black and maroon hyperlinks were neither bold nor underlined. In both versions, global links were presented in Arial font with point size 16 and other links in Arial font with point size 10. The font size of normal text in both versions was Arial 10 point. The navigation menu at the top of each page was used for global navigation and only included in-text links. The rest of the page (the content area) was used to display content and other navigation and contained a mixture of text and hyperlinks. The background of all intranet pages was white (HTML code: #FFFFFF) and the text color (for information displayed in the content area) was always black.

The experiment ran on personal computers (Intel Pentium,333 MHz, 64 Mb RAM, Microsoft NT4 operating system, 14 inch monitors). The screen dimensions were 800×600 pixels. Contrast and brightness were set to optimal levels.

Participants were asked a series of questions based on information available on the intranet (see Appendix A). Answers to the questions were zero (i.e., answers appeared on the home page itself), one, or two links away from the home page. Four sets of rating questions were also presented to participants. Rating questions were selected that satisfied both of the following two criteria: 1) related to influential theoretical concepts in Human-Computer Interaction (technology acceptance: Davis, 1993; Davis & Wiedenbeck, 2001; disorientation: Otter & Johnson, 2000; Ahuja & Webster, 2001; and flow: Davis & Wiedenbeck, 2001); 2) relevant to users' interaction with Web sites. In the first rating task, participants completed the flow scale (9 items) developed by Davis and Wiedenbeck (2001) as a measure of intrinsic motivation. The second, third, and fourth rating tasks used, respectively, the perceived ease of use (3 items), perceived usefulness (4 items), and disorientation (7 items) scales developed by Ahuja and Webster (2001). All four instruments used 7-point Likert scales (see Appendix B). In addition to psychometric measures, measures of speed and efficiency of task performance were collected.

Participants

There were 81 participants, all undergraduates in their first week at university. None was familiar with the Psychology intranet. Thirty-six students took part in the black link color condition, and 45 took part in the blue link condition. The mean age of participants was 23 years with a standard deviation of 7.8 for those in the black condition, and the mean and standard deviation were 22 years and 6.0 respectively in the blue condition. There were 29 females and 7 males in the black condition, and 38 females and 7 males in the blue condition.

Almost all participants (94% and 98% in the black condition and the blue condition, respectively) had some experience of using the Web. Participants had been using the Web from anywhere between one month to more than a year; however, a majority (81% and 76% in the black condition and the blue condition,

respectively) had been using the Web for more than 1 year. Over two-thirds (75% and 73% in the black condition and the blue condition, respectively) used the Web several times a week or at least once a day.

Participants were assigned to one of the experimental conditions according to the lab class they attended. The experiment was carried out in classrooms with 15 to 20 participants at a time. Participants worked individually.

Procedure

In both color conditions, participants were presented with instructions before completing a practice task, which consisted of a series of three information retrieval questions. They were told that a question would appear at the top of the screen. After reading each question, they had to click on a button labeled "Show web site." the home page of the Psychology intranet site was then displayed in the browser window. Participants were instructed to find the answer to each question using the intranet site; the question remained displayed at the top of the screen while participants browsed (see Figure 1). Once they found the answer they had to click a button labeled "Your answer." A dialog box then appeared into which participants entered their answers (see Figure 1). Once they had typed their answer, they were shown the next question. Participants were instructed to work as quickly and accurately as possible.

When all participants had completed the practice task, any questions were answered before they went on to the main part of the task in which they completed a series of 10 further questions (see Appendix A). After the experimental trials, participants completed the rating questions before answering a number of demographic questions (age, sex, use of the Web) that were also presented on screen. Participants took approximately 45 minutes to complete the experiment. When data collection was complete, participants were thanked and debrief.

Data Collection

Experimental software, written in Microsoft Visual Basic version 6,k used the Web browser control (based on version 5 of the Microsoft Internet Explorer Web browser) built into Visual Basic to display intranet pages, which enabled participants to navigate through the site. The program recorded all participants' computer interaction including the sequence of Web pages followed for each task (which incorporated time spent on each page), answers to questions and the time taken to answer each question, as well as responses to questionnaire items.

Results

Disorientation, perceived ease of use, perceived usefulness, involvement, and control were included in psychometric analysis. In addition, accuracy of task performance was calculated as percentage of correct answers and a behavioral efficiency measure as number of pages loaded.

Factor Analysis

A factor analysis revealed five distinct factors (see Table 1). The five factors explained more than two-thirds of the variance in the questionnaire items. The perceived ease of use, perceived usefulness, and disorientation scales each constituted one factor, whereas the flow scale loaded on two separate factors that were named "involvement" (first four flow items) and "control" (second set of five flow items) by Davis and Wiedenbeck (2001). Overall, the factor solution showed simple structure, with the exception of Flow item 6, which had a cross-loading on the involvement factor.

Reliability

The scales for disorientation, perceived usefulness, and perceived ease of use were internally consistent, Cronbach's alphas = 0.92, 0.95, and 0.88, respectively. With alpha = 0.69, the involvement scale was deemed to be sufficiently reliable for further use. Flow item 6 could not be used as this had a cross-loading in the factor analysis; this was excluded from the control scale. Using four items, the control scale was not reliable, alpha = 0.65; however, excluding Flow items 5 and 7 from the scale increased alpha to 0.76. Overall scores, averaged over the items within each of the scales (but only using Flow items 8 and 9 for control), were calculated and used in subsequent analyses.

Validity

Regarding criterion-related validity ("the magnitude of association between measures of a focal construct and measures of another construct expected to co-vary"; Bagozzi, Davis, & Warshaw, 1992), correlations in the following pairs of measures were expected to be significant and confirmed: perceived ease of use and perceived usefulness, $r = 0.47$, $p < 0.001$; disorientation and number of links followed (for links before correct answers), $r = 0.26$, $p < 0.05$; and control and percentage of correct answers, $r = -0.33$, $p < 0.005$. Regarding discriminant validity ("the level of differentiation between measures of different constructs"; Bagozzi et al., 1992), relative low correlations in the remaining psychometric scales were expected. Indeed, these correlations were < 0.30, with only significant correlations between perceived ease of use and control, $r = 0.29$, $p < 0.01$, and between disorientation and control, $r = 0.24$, $p < 0.05$.

Sensitivity

The correlation of control with experience in using the Web was significant, Spearman's $rho = -0.24$, $p < 0.05$ (a lower control score signifies more control). The t-tests showed no effect of link color for disorientation, perceived usefulness, perceived ease of use, involvement, or control (see Table 2), all $p < .05$. The effect of link color on number of pages visited before an incorrect answer was

Table 1. Factor Analysis of Questionnaire Items
(Experiment 1)

	Factor				
Item	Disorientation	Perceived usefulness	Perceived ease of use	Control	Involvement
DIS5	0.878	0.046	−0.169	0.202	−0.006
DIS6	0.851	0.114	−0.070	0.191	−0.035
DIS1	0.836	−0.196	0.149	−0.078	0.027
DIS2	0.815	−0.001	0.043	−0.050	0.023
DIS7	0.799	0.169	−0.192	0.319	−0.072
DIS3	0.776	0.023	0.058	−0.135	−0.040
DIS4	0.767	0.019	−0.141	0.128	−0.163
PUSF3	0.044	0.950	0.138	0.036	0.030
PUSF2	−0.002	0.943	0.115	0.006	0.100
PUSF4	0.047	0.939	0.168	0.068	0.106
PUSF1	0.040	0.785	0.172	0.000	0.136
PEOU2	−0.051	0.405	0.772	0.221	−0.046
PEOU3	−0.035	0.440	0.755	0.099	−0.008
PEOU1	−0.171	0.151	0.743	0.328	0.165
FLO8	0.182	−0.025	0.106	0.807	0.123
FLO9	0.149	0.037	0.216	0.767	−0.178
FLO7	0.075	0.154	−0.066	0.517	0.065
FLO5	−0.093	−0.090	0.184	0.485	0.122
FLO6	0.367	0.016	0.155	0.357	0.360[a]
FLO2	−0.086	0.172	0.185	−0.032	0.737
FLO3	−0.078	0.049	−0.282	0.174	0.717
FLO1	0.040	0.032	0.268	−0.057	0.695
FLO4	−0.145	0.147	−0.368	0.235	0.625
Variance explained	23.41	18.26	11.17	10.67	9.86

[a]Cross-loading.

Note: Extraction method: principal components; rotation method: varimax. DIS: disorientation; PUSF: perceived usefulness; PEOU: perceived ease of use; FLO: intensity of flow.

Table 2. Outcome Measures in the Experiment (Experiment 1)

Outcome	Psychology intranet—blue link color		Psychology intranet—other link colors[a]	
	M	(*SD*)	*M*	(*SD*)
Percentage of correct responses	78.89	(18.61)	74.72	(21.45)
Average number of visited links (correct responses)	4.92	(2.26)	4.50	(1.71)
Average number of visited links (incorrect responses)	7.36	(7.04)	4.16	(3.96)
Disorientation	3.41	(1.30)	2.97	(1.44)
Perceived usefulness	4.37	(1.56)	4.28	(1.71)
Perceived ease of use	4.24	(1.52)	3.70	(1.80)
Involvement	4.49	(1.25)	4.81	(1.05)
Control	3.54	(1.28)	3.44	(1.55)

[a]Black links for global navigation, maroon links for other navigation (see Figure 1).

significant, $t(79) = 2.44$, $p < 0.05$, $d = 0.55$ (probability without Bonferroni correction), but the effect on percentage of correct responses and number of pages visited before a correct answer was not significant, both $t < 1$ (see Table 2).

To summarize the results of Experiment 1, the factor solution showed a distinct set of factors, scales were reliable, evidence for validity was found, but sensitivity as a function of experimental manipulation of link color was not demonstrated generally. Therefore, Experiment 2 was designed with a focus on the sensitivity of the scales.

EXPERIMENT 2

Method

Experimental Design and Materials

An independent measures design with two experimental conditions was used. Type of intranet support was the independent variable and was chosen to address a

dilemma in the design of educational intranet support, where special-purpose systems may offer easier access to relevant materials than general-purpose, but more powerful, VLEs. The two levels of support were special-purpose, bespoke intranet and general-purpose VLE. The intranet site was the same as the version with black and maroon links in Experiment 1. The VLE was a site constructed with exactly the same information content and structure, but this time implemented using the Blackboard™ system (see Figure 2).

The experiment ran on personal computers (Intel Pentium, 1.8 GHz, 512 MbRAM, Microsoft XP, 17 inch monitors). The screen dimensions were 800 × 600 pixels. Contrast and brightness were set to optimal levels.

The same type of information-retrieval questions were asked as in Experiment 1 (see Appendix A). Five sets of rating questions were also presented to participants. Rating questions were selected that satisfied the same criteria as those set for Experiment 1. The first four sets of questions were the same as in Experiment 1. The fifth set was to measure aesthetic value (Lindgaard & Dudek, 2003; Tractinsky et al., 2000), was originally proposed by Tractinsky et al. (2000) and had been used previously to evaluate Web sites (van Schaik & Ling, 2003b). All five instruments used 7-point Likert scales (see Appendix B). In addition to
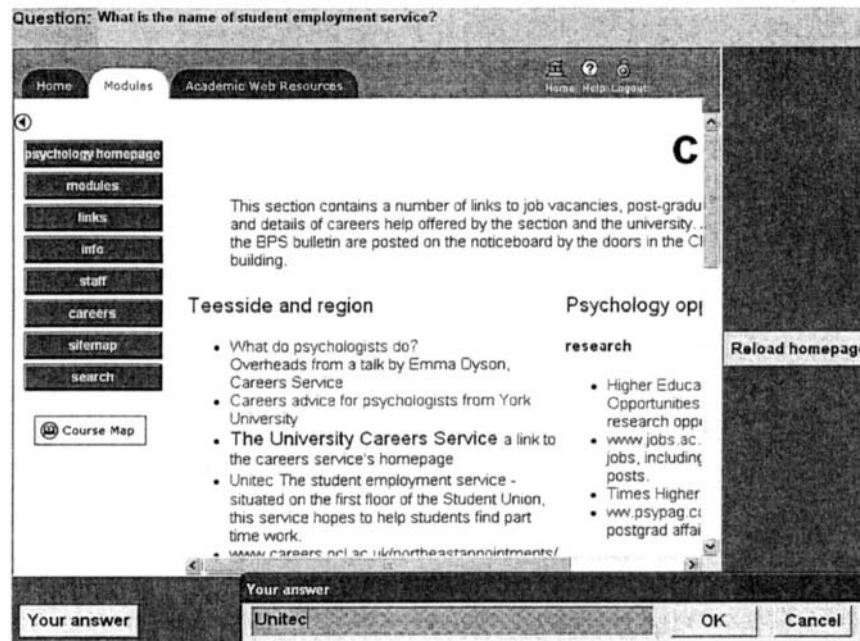

Figure 2. A typical Web page used in Experiment 2 (VLE interface).

psychometric measures, measures of accuracy and efficiency of task performance were collected.

Participants, Procedure and Data Collection

There were 109 participants, all undergraduates in their first week at university. None was familiar with the intranet. Fifty-two students took part in the VLE condition and 57 took part in the intranet condition. Participants' mean age was 21 and 20 years with a standard deviation of 5.1 and 8.1 in the VLE condition and the intranet condition respectively. There were 37 females and 15 males in the VLE condition and 44 females and 14 males in the intranet condition.

Almost all participants (98% in both conditions) had some experience of using the Web. Participants had been using the Web from anywhere between one month to more than a year; however, a majority (81% and 88% in the VLE condition and the intranet condition respectively) had been using the Web for more than 1 year. Over three-quarters (79% and 83% in the LVE condition and intranet condition respectively) used the Web several times a week or at least once a day. Participants were assigned to one of the experimental conditions according to the lab class they attended. The experiment was carried out in classrooms with 15 to 20 participants at a time. Participants worked individually. The same procedure and method of data collection were used as in Experiment 1.

Results

Disorientation, perceived ease of use, perceived usefulness, aesthetic value, involvement, and control were included in psychometric analysis. In addition, accuracy of task performance was calculated as percentage of corrected answers and a behavioral efficiency measure as number of pages loaded.

Factor Analysis

A factor analysis revealed five distinct factors[1] (see Table 3). The five factors explained just under two-thirds of the variance in the questionnaire items. The perceived ease of use, perceived usefulness, disorientation, involvement (first four items of the flow scale; Davis and Wiedenbeck, 2001) and aesthetic scales each constituted one factor.

Reliability

The scales for disorientation, perceived usefulness, perceived ease of use, and aesthetic value were internally consistent, Cronbach's alphas = 0.89, 0.94,

---

[1] A factor analysis extracting six factors failed to show a clear factor structure, with the control items (second set of five items of the flow scale; Davis & Wiedenbeck, 2001) loading on several factors. the control items were excluded from all subsequent analyses.

Table 3. Factor Analysis of Questionnaire Items
(Experiment 2)

	Factor				
Item	Disorientation	Perceived usefulness	Perceived aesthetic value	Involvement	Perceived ease of use
DIS5	0.853	−0.057	−0.162	−0.155	0.055
DIS7	0.825	−0.006	−0.155	−0.151	−0.013
DIS6	0.812	−0.107	−0.264	−0.141	0.002
DIS2	0.772	0.078	−0.084	−0.037	−0.114
DIS1	0.714	−0.051	−0.060	−0.030	−0.050
DIS4	0.620	−0.208	−0.317	−0.374	0.291
DIS3	0.504	−0.311	−0.216	−0.426	0.206
PUSF2	−0.099	0.910	0.047	0.007	0.161
PUSF3	−0.027	0.902	−0.014	0.102	0.272
PUSF4	−0.063	0.864	−0.056	0.105	0.192
PUSF1	−0.017	0.861	0.043	0.141	0.210
AESTH2	−0.065	0.071	0.777	0.030	−0.039
AESTH6	−0.190	0.112	0.745	0.123	−0.258
AESTH3	−0.165	0.030	0.693	0.062	−0.012
AESTH7	−0.291	−0.086	0.660	0.286	0.003
AESTH1	−0.407	0.036	0.649	−0.026	−0.184
AESTH5	−0.060	−0.172	0.645	0.245	0.079
AESTH4	0.067	−0.306	0.322	−0.219	0.232
FLO4	−0.053	0.067	0.347	0.757	−0.051
FLO3	−0.130	0.074	0.129	0.745	0.032
FLO2	−0.292	0.075	0.061	0.572	−0.036
FLO1	−0.038	0.077	−0.022	0.475	0.185
PEOU2	−0.066	0.377	−0.073	0.093	0.774
PEOU3	−0.051	0.367	0.70	0.132	0.754
PEOU1	0.063	0.182	−0.203	−0.027	0.715
Variance explained	17.07	15.09	13.97	9.40	8.83

Note: Extraction method: principal components; rotation method: varimax. DIS: disorientation; PUSF: perceived usefulness; PEOU: perceived ease of use; FLO: intensity of flow.

0.80 and 0.80 respectively. With alpha = 0.64, the involvement scale was deemed to be sufficiently reliable for further use.

Validity

Regarding criterion-related validity, correlations in the following pairs of measures were expected to be significant and confirmed: perceived ease of use and perceived usefulness, $r = 0.51$, $p < 0.001$; and disorientation and number of links followed (for links before correct answers), $r = 0.29$, $p < 0.05$. Regarding discriminant validity, relative low correlations in the remaining psychometric scales were expected. Most of the correlations were 0.30, and significant correlations were found in the following pairs: involvement and disorientation, $r = –0.44$, $p < 0.001$; aesthetic value and disorientation, $r = –0.46$, $p < 0.001$; and involvement and aesthetic value, $r = 0.33$, $p < 0.001$. The first two correlations make sense because disorientation would preclude a feeling of involvement and a positive judgment of aesthetic value of Web pages. The third correlation became non-significant after controlling for disorientation, partial $r = 0.17, p > 0.05$.

Sensitivity

The *t*-tests showed an effect of Web site design on the following scales: disorientation, $t(107) = 3.05$, $p < 0.005$, $d = 0.58$; perceived usefulness, $t(107) = –2.85$, $p < 0.01$, $d = –0.55$; involvement, $t(107) = –2.78$, $p < 0.01$, $d = –0.53$; and aesthetic value, $t(107) = –2.43$, $p < 0.05$, $d = –0.47$ (see Table 4). Further *t*-tests demonstrated that the effect of design on percentage of correct answers, $t(107) = –4.056$, $p < 0.001$, $d = –0.78$, but not on efficiency of navigation (see Table 4).

To summarize the results of Experiment 2, as in Experiment 1 the factor solution showed a distinct set of factors, scales were mostly reliable, evidence for validity was found, but this time sensitivity was also demonstrated as a function of Web site design.

GENERAL DISCUSSION

Psychometric Properties of Subjective Measures

In two experiments, the current study found evidence for the psychometric properties of five scales for the evaluation of educational intranet sites. *Factor analysis* confirmed and extended the findings of previous research (Ahuja & Webster, 2001; van Schaik & Ling, 2003a) by revealing the following five distinct factors: perceived ease of use, perceived usefulness, disorientation, aesthetic value, and involvement (one aspect of flow). These results are very encouraging because—although these factors have been identified in previous research (for

Table 4. Outcome Measures in the Experiment (Experiment 2)

Outcome	Psychology intranet		VLE	
	M	(SD)	M	(SD)
Percentage of correct responses	74.04	(14.98)	59.04	(23.12)
Average number of visited links (correct responses)	4.61	(1.98)	5.44[a]	(3.93)
Average number of visited links (incorrect responses)	10.52	(7.96)	8.34[a]	(5.24)
Disorientation	3.22	(1.15)	3.92	(1.25)
Perceived usefulness	4.39	(1.57)	3.61	(1.29)
Perceived ease of use	3.96	(1.38)	3.94	(1.39)
Involvement	4.63	(0.86)	4.10	(1.11)
Aesthetic value	4.88	(0.69)	4.48	(1.03)

[a]Correction made for two extra page visits required to navigate to answer pages in the VLE compared to the Psychology intranet.

example Ahuja & Webster, 2001; Davis & Wiedenbeck, 2001)— these studies: a) used questionnaire items that represented only subsets of the set of five factors identified in the current study; b) did not administer questionnaires online; and c) did not involve educational intranets. In the current study, the five theoretical constructs were represented as separate factors in a 5-factor structure, based on the data that were collected. However, the items of the control scale (one aspect of flow) only emerged as a factor in Experiment 1 and not in Experiment 2.

When analyzing the *reliability* of the highly loading items for each factor as a corresponding psychometric scale, four scales (perceived ease of use, perceived usefulness, disorientation, and aesthetic value) were highly reliable and the fifth (involvement) was marginally so. These results are also particularly encouraging as the second and fifth scales were originally not developed for use with Web sites. Furthermore, these results are congruent with those of van Schaik and Ling (2003a) who found high reliabilities for the perceived ease of use and disorientation scales (both alpha > 0.85), and the involvement scale (alpha > 0.80) in a study of the use of an on-line shopping site.

In Experiment 1, evidence for the *validity* of the scales was found through significant correlations of perceived ease of use with both perceived usefulness, control with correctness of answers, and disorientation with number of links before correct answers. Further evidence in Experiment 2 demonstrated significant correlations of perceived ease of use with perceived usefulness, and disorientation with page visits. These results are encouraging, in particular for the disorientation scale with evidence for validity in both experiments, in line with results obtained by Ahuja 7 Webster (2001). The validity of perceived usefulness and perceived ease of use (see, e.g., Davis 1993; Davis & Venkatesh, 1996) was also confirmed, but now: a) in the context of educational intranet use; and b) with online questionnaire administration.

In Experiment 1, the control scale was sensitive to differences in experience in Web use. However, none of the psychometric measures exhibited *sensitivity* to differences in link color in Experiment 1. Experiment 2 did establish the sensitivity of the following scales to differences in Web site design: disorientation, perceived usefulness, involvement, and aesthetic value, with medium-range effect sizes ($d = 0.5$; Cohen, 1988). In combination, the two experiments demonstrate that, given the right conditions, scales will exhibit sensitivity. Although the perceived ease of use scale was not sensitive to web site design in Experiment 2, disorientation—which can be seen as a more specific lack of ease of use—was sensitive. These results are encouraging because perceived aesthetic value, involvement, and perceived usefulness have not previously been shown to be sensitive to Web site designs. For the disorientation scale, evidence for sensitivity found by Ahuja and Webster (2001) was confirmed by the results from Experiment 2.

Implications for the Design of Educational Intranets

The sensitivity of measures to a difference in intranet support in Experiment 2 (special-purpose, bespoke Psychology intranet facility versus a general-purpose, "off-the-shelf" VLE) illustrates the more general dilemma in the design of educational intranet support. Special-purpose systems may offer easier access to relevant materials than general-purpose, but more powerful, VLEs. In this case, the bespoke system provided quick access to resources with use of the full computer screen, more straightforward navigation structure and less cluttered screen layout, and with relatively less disorientation and more involvement, perceived ease of use, and perceived usefulness. On the other hand, the VLE provided access to the same materials with a smaller screen area for information display, less straightforward structure and more clutter, and with worse psychometric outcomes. The choice for type of intranet support will depend on various factors, such as levels of skill—of both students and staff—in the use of information technology and Web-based systems in particular, and the extent of desirability to use additional functionality offered by VLEs, such as a synchronic

computer dialogue ("forums") (Webb, Jones, Barker, & van Schaik, 2004) and tracking facilities.

Implications for the Evaluation of Educational Intranets

Our findings have several implications for the evaluation of educational intranets, including VLEs, and Web sites more generally. Table 5 shows the quality of the measures of disorientation, perceived ease-of-use, perceived useful-ness, involvement, and aesthetic value. All of these can be recommended in terms of most of the psychometric properties. An alternative disorientation scale was developed by Lee and Tedder (2004); however, this instrument has not been subjected to factor analysis and its validity is unknown.

The need for the evaluation of educational intranet sites, using good evaluation measures, becomes increasingly important as increasing amounts of learning and teaching materials are made available to students online. This need becomes even more pressing, considering the fact that these materials are ever more exclusively "published" online without hard copies for students. An intranet or VLE becomes the interface to all materials for students and consequently usability, and the quality of human-computer interaction more generally, becomes of para-mount importance for progression and retention of students. Continual evaluation is therefore required to establish how well this interface supports students and where improvements are necessary. In particular, the information architecture of Web sites in general (Rosenfeld & Morville, 2002) and intranets in particular (Nielsen, 2005) will be crucial. For example, lectures have, in principle, an infinite number of ways to structure their learning materials within a VLE in terms of information scent (the extent to which a link is predictive of the content of the destination page), and depth and breadth of information hierarchy (Larson & Czerwinski, 1998). Evaluation of such structures is essential to optimize information architectures.

Table 5. Psychometric Properties of Scales

	Factor structure	Reliability	Validity	Sensitivity
Disorientation	✔	✔	✔	✔
Perceived ease of use	✔	✔	✔	
Perceived usefulness	✔	✔	✔	✔
Aesthetic value	✔	✔	✔	✔
Involvement	✔	(✔)	✔	✔

Note: Properties based on results from Experiments 1 and 2. Ticks denote good psycho-metric properties.

Future Research

Although evidence for validity and sensitivity of the control scale was found in Experiment 1, the scale emerged as an ill-defined factor in Experiment 2. Therefore, further work on the definition of the items of the scale is required. Although the involvement scale formed a clearly defined factor and had good validity and sensitivity, improvement in the formulation of scale items may improve its reliability. Future research should also establish conditions under which the sensitivity of the perceived ease of use scale can be demonstrated.

In conclusion, psychometric properties of five scales for measuring the quality of interaction with intranet sites were confirmed and their use is recommended for evaluating educational intranets. Further research in this area is required to develop a comprehensive suite of proven theory-based psychometric measures to complement performance and behavioral measures.

APPENDIX A
Information Retrieval Questions Presented to Participants

	Information retrieval question	Used in experiment
1	Using the campus map, on which road is the Centuria building located?	1
1	What is the hand-in date of the assessment for the module Research methods 2?	2
2	Where is the Drop in Study Skills Centre (DISSC) located?	1, 2
3	Name one of the two organizations that offer ethical guidelines.	1, 2
4	What is PsycINFO?	1, 2
5	Who should you contact to report any problems with the Psychology intranet site?	1, 2
6	What is the name of the LRC library catalogue?	1
6	Which author's name is presented as an example in the library catalog?	2
7	What is the telephone number of Dr. Dave Woodhouse?	1, 2
8	What is the name of student employment service?	1, 2
9	What is the telephone number of the LRC?	1
9	What is the e-mail address to find information about using library Web site?	2
10	Who deals with requests for coursework extensions?	1, 2

APPENDIX B
Questionnaire Items

Perceived ease of use
Learning to use this site was easy
Becoming skillful at using the site was easy
The site was easy to navigate

Perceived usefulness
Using the site would improve my performance in my coursework
Using the site in my coursework would increase my productivity
Using the site would enhance my effectiveness in my coursework
I would find the site useful in my coursework

Disorientation
I felt lost
I felt like I was going around in circles
It was difficult to find a page that I had previously viewed
Navigating between pages was a problem
I didn't know how to get to my desired location
I felt disoriented
After browsing for a while I had no idea where to go next

Intensity of flow—involvement
I thought about other things
I had to make an effort to keep my mind on the activity
I was aware of distractions
I was aware of other problems

Intensity of flow—control
Time seemed to pass more quickly
I knew the right things to do
I felt like I received a lot of direct feedback
I felt in control of myself
I felt in harmony with the environment

Aesthetic quality
I judge the Web page to be: very complex 1 2 3 4 5 6 7 very simple
I judge the Web page to be: very illegible 1 2 3 4 5 6 7 very legible
I judge the Web page to be: very disordered 1 2 3 4 5 6 7 very ordered
I judge the Web page to be: very ugly 1 2 3 4 5 6 7 very beautiful
I judge the Web page to be: very meaningless 1 2 3 4 5 6 7 very meaningful
I judge the Web page to be: very incomprehensible 1 2 3 4 5 6 7 very
   comprehensible
I judge the Web page to be: very bad 1 2 3 4 5 6 7 very good

Note: Items were presented horizontally using 7-point Likert scale format. Scale end points were "Strongly agree" (on the left0 and "Strongly disagree" (on the right), except for the disorientation items with end points "Never" (on the left) and "Always" (on the right).

REFERENCES

Ahuja, J., & Webster, J. (2001). Perceived disorientation: An examination of a new measure to assess web design effectiveness. *Interacting with Computers*, *14*, 5-29.

Bagozzi, R. P., Davis, F. D., & Warshaw, P. R. (1992). Development and test of a theory of technological learning and usage. *Human Relations*, *45*, 659-686.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. Jordan, B. Thomas, B. Weerdmeester, & I. McClelland (Eds.), *Usability evaluation in industry* (pp. 189-194). London: Taylor & Francis.

Browning, P., & Williams, J. (1997). Using the Internet in teaching and learning: A U.K. perspective. *Computers & Geosciences*, *23*, 549-557.

Campbell, C. S., & Maglio, P. P. (1999). Facilitating navigation in information spaces: Road-signs on the World Wide Web. *International Journal of Human-Computer Studies*, *50*, 309-327.

Chen, H., Wigand, R., & Nilan, M. (1999). Optimal experience of web activities. *Computers in Human Behavior*, *15*, 585-608.

Chin, J. P., Diehl, V. A., & Norman, K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Conference on human factors in computing systems* (pp. 213-218). New York: Association for Computing Machinery.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conklin, J. (1987). Hypertext: An introduction and survey. *IEEE Computer*, *20*, 17-41.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly/September*, *13*, 318-340.

Davis, F. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Human-Computer Studies*, *45*, 19-45.

Davis, S., & Wiedenbeck, S. (2001). The mediating effects of intrinsic motivation, ease of use and usefulness perceptions on performance in first-time and subsequent computer users. *Interacting with Computers*, *13*, 549-580.

Kirakowski, J. (1996). The software usability measurement inventory: Background and usage. In P. Jordan, B. Thomas, B. Weerdmeester, & I. McClelland (Eds.), *Usability evaluation in industry* (pp. 169-177). London: Taylor & Francis.

Kirakowski, J., & Cierlik, B. (1998). Measuring the usability of web sites. In *Proceedings of the 42nd Human Factors and Ergonomics Society annual conference* (pp. 424-428). Santa Monica, CA: Human Factors and Ergonomics Society.

Larson, K., & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent for information retrieval. *Proceedings of CHI 1998* (pp. 25-32). Los Angeles, CA: Association for Computing Machinery.

Lee, M., & Tedder, M. (2004). Introducing expanding hypertext based on working memory capacity and the feeling of disorientation: Tailored communication through effective hypertext design. *Journal of Educational Computing Research*, *30*, 171-195.

Lewis, J. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability testing. *International Journal of Human-Computer Interaction*, *14*, 463-388.

Lindgaard, G., & Dudek, C. (2003). What is this evasive beast called user satisfaction? *Interacting with Computers*, *15*, 429-452.

Moneta, G., & Csikszentmihalyi, M. (1996). The effect of perceived challenges and skills on the quality of subjective experience. *Journal of Personality*, *64*, 275-310.

Murgolo-Poore, M., Pitt, L., & Ewing, M. (2002). Intranet effectiveness: A public relations paper-and-pencil checklist. *Public Relations Review*, *28*, 113-123.

Nielsen, J. (1999). *Intranet portals: The corporate information infrastructure*. [Column posted on the World Wide Web.] Mountain View, CA: Author. Retrieved August 25, 2004 from the World Wide Web: http://www.useit.com/alertbox/990404.html

Nielsen, J. (2001). *The 10 best intranet designs of 2001*. [Column posted on the World Wide Web.] Mountain View, CA: Author. Retrieved August 25, 2004 from the World Wide Web: http://www.useit.com/alertbox/20011125.html

Nielsen, J. (2005). *The canonical intranet homepage*. [Column posted on the World Wide Web.] Mountain View, CA: Author. Retrieved May 24, 2005 from the World Wide Web: http://www.useit.com/alertbox/20050523.html

Otter, M., & Johnson, H. (2000). Lost in hyperspace: Metrics and mental models. *Interacting with Computers*, *13*, 1-40.

Rosenfeld, L., & Morville, P. (2002). *Information architecture for the World Wide Web* (2nd ed.). Sebastopol, CA: O'Reilly.

Stalidis, G., Prentza, A., Vlachos, I., Maglavera, S., & Koutsouris, D. (2001). Medical support system for continuation of care based on XML web technology. *International Journal of Medical Informatics*, *64*, 385-400.

Thong, J., Hong, W., & Tam, K.-Y. (20902). Understanding user acceptance of digital libraries: What are the roles of interface characteristics, organizational context, and individual differences? *International Journal of Human-Computer Studies*, *59*, 545-567.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*, 127-145.

van Schaik, P., & Ling, J. (2003a). Using online surveys to measure three key constructs of the quality of human-computer interaction in Web sites: Psychometric properties and implications. *International Journal of Human-Computer Studies*, *59*, 545-567.

van Schaik, P., & Ling, J. (2003b). The effect of link colour on information retrieval in educational intranet use. *Computers in Human Behavior*, *19*, 553-564.

Webb, E., Jones, A., Barker, P., & van Schaik, P. (2004). Towards a classification of e-dialogue learning characteristics in students in Higher Education. *Innovations in Education and Teaching International*, *41*, 93-103.

Direct reprint requests to:

Dr. Paul van Schaik
Psychology Section
School of Social Sciences and Law
University of Teesside
Middlesbrough, TS1 3BA, U.K.
e-mail: p.van-schaik@tees.ac.uk