

Web Site Evaluation: Towards a Validated Instrument

Mike Lopez
Christchurch Polytechnic
Institute of Technology
130 Madras Street,
Christchurch, New Zealand
+64 3 940 8000
mike.lopez@cpit.ac.nz

David Weir
Christchurch Polytechnic
Institute of Technology
130 Madras Street,
Christchurch, New Zealand
+64 3 940 8000
david.weir@cpit.ac.nz

Luofeng Xu
Christchurch Polytechnic
Institute of Technology
130 Madras Street,
Christchurch, New Zealand
+64 3 940 8000
luofeng.xu@cpit.ac.nz

ABSTRACT

In this poster paper, we describe the development and initial validation of an instrument for Web Site Evaluation. The instrument is a questionnaire with 61 questions organised in seven scales. The instrument was trialled with 33 level five computing students. Overall, the instrument was found to be reliable. Four of the scales demonstrated acceptable coherence. However three scales require further work.

Categories and Subject Descriptors

K.3.2 [Computers and Education]: Computer and Information Science Education

General Terms

Measurement, Performance, Reliability, Standardisation.

Keywords

Web site evaluation, technology acceptance.

1. INTRODUCTION

Our project had its origins in work carried out by a third year computing student in which the student developed an evaluation tool for web sites [Jason Hsiao, personal communication, 2012]. Using this as a starting point, we created a pilot instrument that was more solidly grounded in theory and in extant standards such as the W3C accessibility guidelines. Our instrument has seven major scales.

We tested the instrument in a level five course in a three year computing degree. Students were asked to use the tool to evaluate three major Web sites (Microsoft, Vodafone and Snap).

We then used their evaluations to analyse the psychometric properties of the instrument. The instrument appears to have strong psychometric properties overall, but issues were found with some subscales. We plan to trial the instrument in other contexts and with a wider range of web sites. We also plan to develop the instrument further by refining the questions asked and creating simplified and extended versions.

Once the instrument is validated we will use it to inform the rubrics we use for web site assignments. The instrument may also have a wider use in industry. We plan to use the instrument as a research tool to investigate cultural and gender differences in perceptions of web sites. Finally, a key goal in creating the instrument was to encourage our students to think critically about Web sites and to help them develop professional judgement about what makes a Web site effective. We plan to evaluate how well this goal has been achieved.

There are several parts to this project. This poster describes the initial validation of the pilot instrument.

2. THEORETICAL FRAMEWORK

The technology acceptance model [1] provides a useful conceptual framework that explains acceptance of technology. It identifies two key components: perceived usefulness (PU), and perceived ease of use (PEU). However, the TAM assumes that technology is used to achieve a useful task. On the Web, appearance is also an important factor so we have extended TAM to include this.

The Web Accessibility Initiative (WAI) from the W3C gives standards and guidelines for accessibility [4]. We used these as the basis for the accessibility scale.

There are numerous instruments that focus on usability. Tullis and Stetson give a summary of five instruments [3]. Our approach was to bring together these various perspectives into a single comprehensive instrument,

**Web Site Evaluation:
Towards a Validated Instrument**

Dr Michael Lopez
Christchurch Polytechnic
Institute of Technology
mike.lopez@cpit.ac.nz

Dr David Weir
Christchurch Polytechnic
Institute of Technology
david.weir@cpit.ac.nz

Dr Luofeng Xu
Christchurch Polytechnic
Institute of Technology
luofeng.xu@cpit.ac.nz

The project

Our project had its origins in work carried out by a third year student in the BECT in which the student developed an evaluation tool for web sites (Jason Hsiao, personal communication, 2012).

Using this as a starting point, we created a pilot instrument that was more solidly grounded in both theory and extant standards such as the W3C accessibility guidelines.

Our instrument has seven major scales. We tested the instrument with 33 students who were asked to use the tool to evaluate three major web sites (Microsoft, Vodafone and Snap).

We then used their evaluations to analyse the psychometric properties of the instrument. The instrument appears to have strong psychometric properties overall, but issues were found with some subscales.

We plan to trial the instrument in other contexts and with a wider range of web sites. We also plan to develop the instrument further by refining the questions asked and creating simplified and extended versions.

Once the instrument is validated we will use it to inform the rubrics we use for Web site assignments. The instrument may also have a wider use in industry.

We plan to use the instrument as a research tool to investigate cultural and gender differences in perceptions of web sites.

Finally, a key goal in creating the instrument was to encourage our students to think critically about Web sites and to help them develop professional judgement about what makes a Web site effective. We plan to evaluate how well this goal has been achieved.

Instrument and Validation

The pilot instrument is a questionnaire with 61 questions organised into seven scales. Responses are on a five-point Likert scale.

We tested the instrument with 33 evaluations from 33 students. No instructions were given to students on how to carry out the evaluations or interpret the questionnaire forms.

Initial indications are that the pilot instrument has strong overall psychometric properties. Reliability, as measured by Cronbach's alpha is shown in the table to the right. Some of the items were diagnosed as degrading and an improved version was derived by cutting these items. The reliability of the improved instrument is shown in the same table.

The content, appearance, navigation and usability subscales seem acceptable coherent, with alpha above .8. The remaining three sub-scales need improvement.

Work on construct validity, discriminant validity, convergent validity and factor analysis of the results is on-going. We need more data before we can confidently report on these.

Theoretical Framework

The technology acceptance model (Davis, 1989) provides a useful conceptual framework that explains acceptance of technology. It identifies two key components: perceived usefulness (PU), and perceived ease of use (PEU). However, the TAM assumes that technology is used to achieve a useful task. On the Web, appearance is also an important factor so we have extended TAM to include this.

The Web Accessibility Initiative (WAI) from the W3C gives standards and guidelines for accessibility [4]. We used these as the basis for the accessibility scale.

There are numerous instruments that focus on usability. Tullis & Stetson (2004) give a summary of five instruments.

Reliability

	Original	Improved		
Scale	N	Alpha	N	Alpha
Content	9	.713	7	.887
Appearance	12	.854	11	.866
Navigation	9	.862	7	.885
Usability	11	.760	8	.825
Accessibility	12	.722	11	.724
Performance	3	.376	3	.376
Consistency	5	.722	5	.727
Overall	61	.931	52	.946

Note: Improved alpha values displayed in bold.

Conclusion

This early pilot stage achieved strong overall reliability and satisfactory reliability in four of seven sub-scales.

This gives us confidence that we will be able to develop a reliable instrument for Web site evaluation.

Future Work

We would like to trial and evaluate the instrument in a wider set of contexts. We invite those interested in trialling the instrument to contact the authors at the email addresses shown at the top of this poster.

CPIT

This poster paper appeared at the 4th annual conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2013) incorporating the 26th Annual Conference of the National Advisory Committee on Computing Qualifications, Hamilton, New Zealand, October 6-9, 2013. Mike Lopez and Michael Verhaart, (Eds).

3. METHOD

The pilot instrument is a questionnaire with 61 questions organised into seven scales. Responses are on a five-point Likert scale. We tested the instrument with 86 evaluations from 33 students. No instructions were given to students on how to carry out the evaluations or interpret the questionnaire items.

To validate the instrument, we fitted the data to a linear measurement model [2]. Model fit was assessed with the Rasch infit and outfit statistics [6]. Following guidelines given by Wright and Linacre [5], we classified an item as *degrading* if either the infit or outfit statistics exceeded 2.0.

We calculated reliability both for the original instrument and for an improved instrument in which we removed the items that were classified as degrading.

4. FINDINGS

We use Cronbach's alpha to report the reliability of the instrument and scales. The estimated reliabilities of both the original and improved scales are shown in Table 1.

Table 1: Reliabilities of Scales

Scale	Original		Improved	
	N	Alpha	N	Alpha
Content	9	.773	7	.887
Appearance	12	.854	11	.866
Navigation	9	.862	7	.885
Usability	11	.760	8	.825
Accessibility	12	.722	11	.724
Performance	3	.376	3	.376
Compatibility	5	.727	5	.727
Overall	61	.937	52	.946

Note: N is the number of items. The improved scale is after culling degrading items.

5. DISCUSSION

Initial indications are that the pilot instrument has strong overall psychometric properties. The content, appearance, navigation and usability subscales seem acceptably coherent, with alpha above .8. The remaining three sub-scales need improvement.

Work on construct validity, discriminant validity, convergent validity and factor analysis of the results is on-going. However, we need more data before we can confidently report on these.

6. CONCLUSION

This early pilot stage achieved strong overall reliability and satisfactory reliability in four of seven sub-scales. This gives us confidence that we will be able to develop a comprehensive reliable instrument for Web site evaluation.

7. REFERENCES

- 1 Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 13, 3, 319–340.
- 2 Lopez, M. 2012. Measurement of challenge and self-efficacy in learning [Doctoral Thesis], Perth: Curtin University.
- 3 Tullis, S., & Stetson, J. N. 2004.. A comparison of questionnaires for assessing website usability. Boston, MA. Retrieved from <http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf>
- 4 W3C. n.d. Web Accessibility Initiative (WAI). Retrieved 9 5, 2013, from W3C standards: <http://www.w3.org/standards/webdesign/accessibility>
- 5 Wright, B & Linacre, J. 1994. Reasonable mean-square fit values, *Rasch Measurement Transactions*, 8, 3, 370.
- 6 Wright, B & Stone, M. 1999. Measurement essentials, 2nd. ed., Wilmington, Delaware: Wide Range inc..