The System Usability Scale and Non-Native English Speakers

Kraig Finstad

Intel® Corporation

1385 NW AmberGlen Pkwy

Beaverton, OR 97006

kraig.a.finstad@intel.com

Abstract

The System Usability Scale (SUS) was administered verbally to native English and non-native English speakers for several internally deployed applications. It was found that a significant proportion of non-native English speakers failed to understand the word "cumbersome" in Item 8 of the SUS (that is, "I found the system to be very cumbersome to use.") This finding has implications for reliability and validity when the questionnaire is distributed electronically in multinational usability efforts.

Keywords

Usability findings, usability methods, survey, questionnaire, language, multinational, multicultural, System Usability Scale, international usability, SUS

Introduction

The System Usability Scale (SUS) is a commonly used, freely distributed, and reliable questionnaire consisting of 10 items. Scoring the questionnaire yields a usability score in the range of 0–100. Its ease of administration and scoring makes it a popular choice among usability professionals. Recent research (Tullis & Stetson, 2004) has shown the SUS to provide superior assessments of website usability, compared to other questionnaires (for example, QUIS, CSUQ). Additionally, it has been found

to have a 0.8588 correlation with the 50-item SUMI questionnaire (Holyer, 1993).

The SUS initially was developed by John Brooke while he was at Digital Equipment Corporation (1996). The SUS uses a five-point scale with anchors for Strongly Agree and Strongly Disagree. It is reproduced below from Brooke (1996).

Figure 1. The System Usability Scale

In addition to being a popular choice for online usability surveys, the SUS can be used as a follow-up subjective measure after testing the usability of functional systems as a component of pre- and post-testing. Such verbal administration of the SUS has provided opportunities to gather feedback on the specifics of the questionnaire itself. Observation through verbal administration of the SUS during usability testing has provided several insights into the questionnaire items.

For example, the use of "very" in Items 7, 8, and 9 might be problematic in that they may conflict with the anchor labeling, "strongly." A test participant expressed difficulty agreeing to Item 9, because any measure of agreement would automatically categorize him as "very confident", when he was more comfortable with agreeing he was simply "confident."

It was more consistently noticed that some English-speaking usability test participants in Europe and Asia did not understand what the word "cumbersome" meant in Item 8, "I found the system very cumbersome to use." However, they did tend to understand the word "awkward" when the test facilitator offered it in assistance. Lack of comprehension in such a questionnaire item could seriously compromise the validity of the scale if administered electronically and without the support of a facilitator.

This study was designed to determine how widespread this comprehension problem was, coinciding with an effort aimed at establishing the usability baselines of current systems.

Method and Process
Participants
Participants were 36 Intel® Corporation employees who were categorized into one of two conditions: native English speaking and non-native English speaking. The

native languages of the participants who fit the latter condition were Russian, German, Mandarin Chinese, Filipino, Spanish, and Hebrew. Each condition consisted of 18 participants.

Measures

The critical measure for the study was comprehension of Item 8 of the SUS (i.e., "I found the system very cumbersome to use."). At that point, if the participant voluntarily indicated that they did not understand the word "cumbersome", the facilitator assisted by defining it as "awkward." For each participant, the facilitator noted comprehension success or failure of this item.

Procedure

Relevant measures were taken as part of a survey effort aimed at establishing usability baselines for current software solutions. Participants were contacted by phone, and the SUS was shared via Windows® NetMeeting®. Participants read each item of the SUS out loud and then gave a verbal response to the facilitator, who recorded each response. At the end of the questionnaire was a final item asking the participant's native language. Sessions were facilitated by Kraig Finstad and Carie Green, both Intel® employees.

Results

No other items in the questionnaire needed additional definitions, and the facilitator did not inquire about the understanding of any items. Additionally, for those requesting assistance with Item 8, no further clarification was necessary beyond "awkward."

None of the 18 native English-speaking participants requested assistance with Item 8. Of the 18 remaining non-native English-speaking participants, nine (or 50%) requested assistance with the word "cumbersome" in Item 8. The difference between the two groups was found to be statistically significant (Fisher's Exact Test, p = .001). The differences between these groups is illustrated in Figure 2.
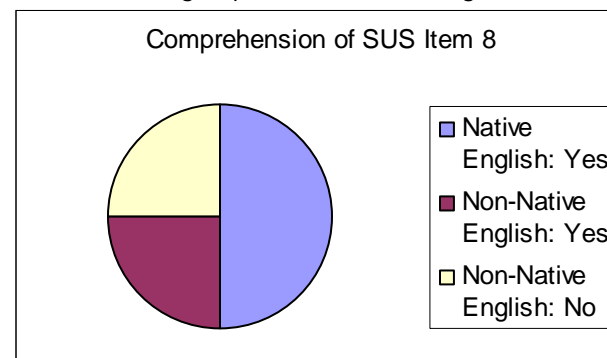

Figure 2. Participant comprehension by native language

Discussion and Recommendations

This finding indicates that the unmodified SUS is likely to be problematic with an international English-speaking audience, especially in an unsupervised medium such as an automated web tool. The problem can be especially pronounced in a composite score tool like the SUS, where the final range of scores depends on all 10 items for a meaningful final calculation. Any missing or incorrectly entered data affects not only the singular questionnaire item, but the scale as a whole. This finding implies that the final score of the SUS in a remote, unsupervised, international usability evaluation might be biased due to a misunderstanding or misinterpretation of terminology in the scale.

An immediate and simple way to prevent such issues is to reword Item 8 as "I found the system very cumbersome/awkward to use", which clarifies the wording while retaining the original language and, as such, should not impact the scale's validity. No users in this study expressed any confusion about the term "awkward", and it captures some of the subtlety of "cumbersome" in a way that other alternatives, like "difficult", do not.

Long-term developments of this research should include a follow-up study contrasting the original SUS with the a modified version that includes the recommendation described above. Additionally, attention should be paid to some of the potentially biasing and/or conflicting language in the scale (i.e., the use of "very" within questionnaire items).

Practitioners' Takeaways

- The System Usability Scale (SUS) as originally conceived may not be suitable for electronic international distribution.
- Many non-native English speakers do not understand the word "cumbersome" in SUS Item 8, but do understand "awkward".
- A simple fix for SUS Item 8 is to substitute "cumbersome" with "cumbersome/awkward."

References
Brooke, J. (1996) SUS: a "quick and dirty" usability scale. In P W Jordan, B Thomas, B A Weerdmeester & A L McClelland (eds.) Usability Evaluation in Industry. London: Taylor and Francis.

Holyer, A. (1993). Methods for Evaluating User Interfaces. Cognitive Science
Research Paper No. 301, School of Cognitive and Computing Sciences, University of Sussex.

Tullis, T.S., & Stetson, J.N. (2004). A Comparison of Questionnaires for Assessing Website Usability. *Proceedings of UPA 2004* (7-11 June 2004).

Kraig Finstad is a Senior Human Factors Engineer at Intel® Corporation, and is interested in usability metrics and problem-solving. Kraig received his Ph.D. in Experimental Psychology from the University of New Mexico.