

Autoren: Bangor, A., Kortum P.T. & Miller, J.T:

Titel: An Empirical Evaluation of the System Usability Scale

Jahr: 2008

Veröffentlicht in: International Journal of Human-Computer Interaction, 24:6, 574-594

Laufende Nummer im Literaturverzeichnis: SUS [1]

Schlagwörter: System Usability Scale, SUS, Reliabilität, Robustheit, SUS Interpretation

Kurzbeschreibung:

Das Paper evaluiert den System Usability Scale Fragebogen aus dem Jahr 1986 hinsichtlich seiner Reliabilität und der Robustheit gegenüber der Art des Interfaces, Alter und Geschlecht, sowie einer einfachen verbalen Interpretation des Ergebnisses. Dabei wurden 2324 aus 206 unterschiedlichen Studien untersucht.

Um die Reliabilität zu überprüfen, wurde Cronbach's Alpha ermittelt. Mit einem Wert von .911 wurde der SUS als reliabel eingestuft um Usability Konzepte messen zu können.

Zur Überprüfung der Robustheit wurden aus den Studien diejenigen ausgesucht, die in die folgenden Interfaceklassen einzuordnen waren: Handyzubehör, customer premise equipment (Telefone, Modems), GUI's für Computerinterfaces, IVR (interactive voice response), Webbasierte Interfaces. Mittels einer einfaktoriellen Varianzanalyse für die Bewertung einzelner Items wurde gezeigt, dass es keine Unterschiede bei der Bewertung zwischen den verschiedenen Interfaces gab.

Hinsichtlich der Robustheit gegenüber Alter und Geschlecht wurden die Studien in Betracht genommen, die diese demografischen Daten erfasst haben. Mittels eines zweiseitigen t-Tests wurden die SUS-Mittelwerte beider Geschlechter verglichen. Sie unterschieden sich nicht. Hinsichtlich des Zusammenhangs zwischen Alter und SUS-Scores wurde eine signifikante negative Korrelation von $-.203$ gefunden, die zeigt, dass ältere Menschen die Interfaces tendenziell schlechter bewerten. Die Autoren fügten hinzu, dass dieser Zusammenhang jedoch weiterer Untersuchungen bedarf.

Um den SUS-Score nach der Auswertung schnell einstufen zu können, wurde der Standardfragebogen um eine 11. Frage erweitert, die nach einer generellen Einschätzung der Benutzerfreundlichkeit fragt. Die Antwortmöglichkeiten wurden durch eine 7-stufige Adjektivskala gegeben, die von best-imaginable über ok bis worst-imaginable reicht. In der Studiensammlung waren 212 Studien, die diese Frage enthielten. Die Korrelation zwischen dem generellen SUS-Score und der Bewertung auf der Adjektivskala wurde mit $r=.806$ als sehr hoch beschrieben.

Zusammenfassend wurde SUS-Fragebogen als ein valides und robustes Instrument eingestuft. Die Methoden, die die Autoren angewandt haben, sind in den Bereich der grundlegenden Statistik einzuordnen, sodass der Hintergrund der jeweiligen Methode nachvollziehbar war. Aufgrund der großen Stichprobe können die Aussagen als vertrauenswürdig eingestuft werden.

Erstellt von: Josephine Grauert, 346566