

Comparison of German Translations of the System Usability Scale – Which to Take?

Tobias J. BRIX^{a,1}, Alice JANSSEN^a, Michael STORCK^a, and Julian VARGHESE^a

^a*Institute of Medical Informatics, University of Münster, Germany*

Abstract. The System Usability Scale (SUS) is a reliable tool for usability measurement and evaluation. Since its original language is English, a translation is required before a target group can answer it in their native language. The challenge of translating questionnaires lies in the preservation of its original properties. Different versions of a German SUS have been proposed and are currently in use. Objective of this work is to find and compare available German translations. Four versions were found and compared in terms of the translation process and the exact wording of the translation. Only the version of Gao et al. has been systematically validated, but has an unnatural wording. Although not validated yet, the proposed version of Rummel et al. is a good compromise between wording and methodically clean development. The version of Lohmann and Schäffer is the close runner up, as it may improve the wording at the expense of methodological accuracy. Since the version of Rauer gives no information about its translation process, it is considered least preferred of the four compared translations.

Keywords. System Usability Scale, SUS, Questionnaire, Translation, German

1. Introduction

In 1996, John Brooke presented the System Usability Scale (SUS) as a reliable, quick, easy and freely distributed standardized method for subjectively assessing usability [1]. It is used to assess a large number of systems such as mobile applications, websites or expert systems, and is a popular choice for online usability surveys [2].

It is a five-level Likert scale with ten items. The items cover the concepts of effectiveness, efficiency and satisfaction. Participants indicate their agreement with each item on the five-point scale. The items were chosen and ordered in a way that the common response alternates between ‘strongly disagree’ and ‘strongly agree’ to ensure that the respondent reads each statement and thinks about it. The original English wording can be found in the first column of Table 1. The SUS yields a single number between 0 and 100, higher scores being associated with better usability. SUS scores better than 71.4 can be interpreted as ‘good’ [3].

Since the SUS is in English, some problems arise when performing surveys among non-native English speakers. Lack of comprehension, misunderstandings and extended testing time could affect the response score of the SUS [4,5]. Especially, it is not possible to question people who are not proficient in English. Therefore, translations of the SUS exist, for example in French [2], Persian [6], or American Sign Language [7].

¹ Corresponding Author, Tobias J. Brix, Institute of Medical Informatics, University of Münster, Albert-Schweitzer-Campus 1, 48143 Münster, Germany; E-mail: tobias.brix@uni-muenster.de.

Some challenges occur when translating questionnaires. The meaning of the original question items needs to remain unchanged during translation [7]. In addition, psychometric properties like validity, i.e., the questionnaire measures that was intended to be measured, or repeatability need to be preserved [2]. To meet this challenges, researchers apply methods like reverse translation or evaluations through user studies [2,6,7]. Currently, different translations of the SUS in the German language are used to evaluate systems. This makes it difficult to compare the resulting SUS scores reliably.

Objective of this work was to perform a scoping review for available German translations of the SUS and to compare them in terms of exact wording and the methodology used to create them.

2. Methods

A scoping review for German translations of the SUS was performed following the PRISMA guidelines [8]. The used search term was “System Usability Scale” AND “German” within Web of Science (www.webofscience.com), Scopus (www.scopus.com), PubMed (www.pubmed.gov) and dblp (www.dblp.org). The review has been conducted in August 2022, thus all publications until end of July 2022 were considered. As inclusion criterion, translations were chosen if their authors’ goal was to propose a translation to be used for other studies. Translations only created and applied as part of a single study were excluded. These criteria have been validated by two researchers independently. Even if a publication was not selected for comparison, its references were still used in a back and forward search. In addition, already known sources of german translations have been added by the authors. The final comparison focused on the creation process of the translation and the exact wording of the ten items and labeling of the Likert scale.

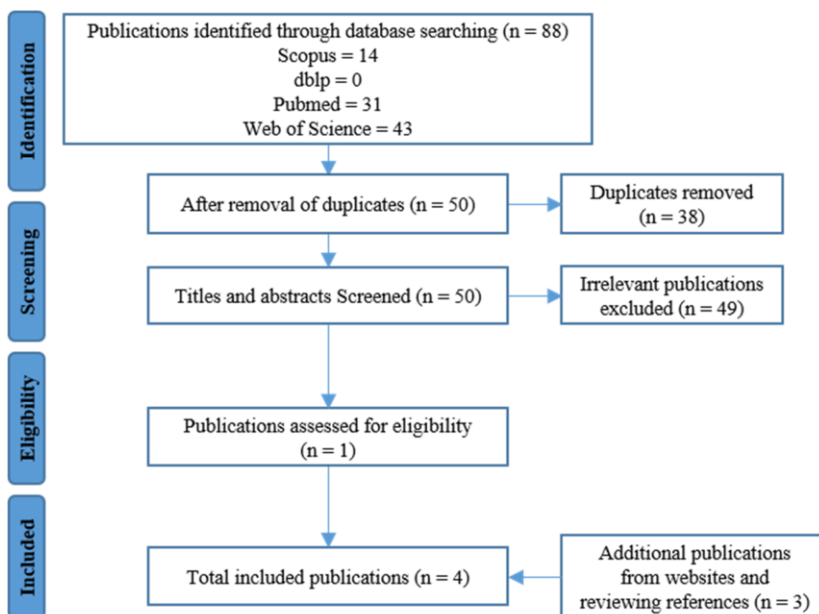


Figure 1. Flow chart of article selection

3. Results

The individual intermediate steps of the PRISMA workflow can be seen in Figure 1. The result of the literature search yielded only a single eligible result [5], two additional translations were found through back and forward search [9,10] and one from a known source [11]. Thus, the comparison included a single peer-reviewed publication [5] and three blog posts [9–11].

The first aspect of comparison was the creation of the translation. Rummel et al. used the input of the crowdsourcing project initiated by Reinhardt [9,12]. Reinhardt started a crowdsourcing project to translate the SUS. The idea was to create a draft from the submissions, reverse translating it into English, and iterate over these two steps until a widely accepted translation of the SUS is reached. The project was never finished. However, Rummel et al. picked up the idea and used the input to propose their translation. The submitting persons are referred to as ‘usability professionals’. Two persons from the SAP Global Design Enablement team consolidated the input. British and US native speakers translated the German translation back to English. Lohmann and Schäffer [10] based their translation on the one by Rummel et al. They applied some changes due to disagreement with the wording. An on-site study with 89 participants was performed. Rauer [11] created his translation about two years before Rummel et al. The blog entry is in German but Rauer does not address the translation process. A professional translation agency translated the version from Gao et al. [5]. The process includes different verification steps, multiple persons, native/proficient speakers and Google translate. They analyzed the study extensively with, for example, Cronbach’s alpha and a factor analysis.

The second aspect of the comparison was the exact wording of the translations. Differences exist especially in the beginning of the items 1 to 8 (‘I think’, ‘I find’, etc.) but also in the articles (‘das System’, ‘dieses System’), the tenses and the word choice. Table 1 shows the results of this comparison.

4. Discussion

Four translations of the SUS into German were compared. In all of them, important information about the creation process is missing and only Gao et al. made a comprehensive validation. Rauer did not describe how or who performed the translation at all. There is also no information about validation. Consequently, it is the less suited translation to use out of this four.

Rummel et al. describes the translation process but questions such as how many translations has the crowdsourcing yielded, how did they proceed to choose the final translation or what is the professional background of the persons who offered translations remains unanswered.

Lohmann and Schäffer adapted the version by Rummel et al., so the initial translation process remains the same. Information about who was involved in the creation of the new version is missing as well. Some insights about the motivation behind the changes would have been very interesting, especially as some of the difference are very small and only the change to item 10 was explained. Furthermore, changing the items proposed by Rummel et al. revoked the result from the reverse-translation. The authors used the translation in an on-site user study, but it was not properly evaluated.

Table 1. The translations of the SUS by Rauer, Rummel et al., Lohmann and Schäffer and Gao et al. [5,9–11] compared to the English version. Cell background: present or past tense. Bold font: the/das or this/dieses. Italic and underline font: different beginning verbs. Grey highlighting: particular wording difference between the translations.

	Brooke [1]	Rauer [11]	Rummel et al. [9]	Lohmann and Schäffer [10]	Gao et al. [5]
1	I <i>think</i> that I would like to use this system frequently.	Ich kann mir sehr gut <u>vorstellen</u> , das System <u>regelmäßig zu nutzen</u> .	Ich <u>denke</u> , dass ich das System gerne häufig benutzen würde.	Ich <u>denke</u> , dass ich dieses System gerne regelmäßig nutzen würde.	Ich <u>denke</u> , dass ich dieses Produkt häufig verwenden möchte.
2	I <u>found</u> the system unnecessarily complex.	Ich <u>empfinde</u> , das System als unnötig komplex.	Ich <u>find</u> das System unnötig komplex.	Ich <u>find</u> das System unnötig komplex.	Ich <u>find</u> das Produkt unnötig komplex.
3	I <u>thought</u> the system was easy to use.	Ich <u>empfinde</u> , das System als einfach zu nutzen.	Ich <u>find</u> das System einfach zu benutzen.	Ich <u>denke</u> , das System war leicht zu benutzen.	Ich <u>denke</u> , das Produkt war einfach zu bedienen.
4	I <u>think</u> that I would need the support of a technical person to be able to use this system.	Ich <u>denke</u> , dass ich technischen Support brauchen würde, um das System zu nutzen.	Ich <u>glaube</u> , ich würde die Hilfe einer technisch versierten Person benötigen, um das System benutzen zu können.	Ich <u>denke</u> , ich würde die Unterstützung einer fachkundigen Person benötigen, um das System benutzen zu können.	Ich <u>denke</u> , dass ich die Unterstützung einer technischen Person brauche, um dieses Produkt nutzen zu können.
5	I <u>found</u> the various functions in this system were well integrated.	Ich <u>finde</u> , dass die verschiedenen Funktionen des Systems gut integriert sind.	Ich <u>find</u> , die verschiedenen Funktionen in diesem System waren gut integriert.	Ich <u>find</u> , die verschiedenen Funktionen des Systems waren gut integriert.	Ich <u>find</u> , die verschiedenen Funktionen in diesem Produkt waren gut integriert.
6	I <u>thought</u> there was too much inconsistency in this system.	Ich <u>finde</u> , dass es im System zu viele Inkonsistenzen gibt.	Ich <u>denke</u> , das System enthält zu viele Inkonsistenzen.	Ich <u>halte</u> das System für zu inkonsistent.	Ich <u>denke</u> , dass dieses Produkt nicht konsistent genug war.
7	I would <i>imagine</i> that most people would learn to use this system very quickly.	Ich kann mir <u>vorstellen</u> , dass die meisten Leute das System schnell zu beherrschen lernen.	Ich kann mir <u>vorstellen</u> , dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.	Ich <u>glaube</u> , dass die meisten Menschen sehr schnell lernen würden, mit dem System umzugehen.	Ich würde mir <u>vorstellen</u> , dass die meisten Leute sehr schnell lernen würden, dieses Produkt zu benutzen.
8	I <u>found</u> the system very cumbersome to use.	Ich <u>empfinde</u> , die Bedienung als sehr umständlich.	Ich <u>find</u> das System sehr umständlich zu nutzen.	Ich <u>find</u> das System sehr umständlich zu benutzen.	Ich <u>find</u> dieses Produkt sehr umständlich zu benutzen.
9	I felt very confident using the system.	Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.	Ich <u>fühle</u> mich bei der Benutzung des Systems sehr sicher.	Ich <u>fühle</u> mich bei der Nutzung des Systems sehr sicher.	Ich habe mich sehr selbstsicher gefühlt, dieses Produkt zu verwenden.
10	I needed to learn a lot of things before I could get going with this system.	Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten konnte.	Ich musste eine Menge lernen, bevor ich anfangen konnte das System zu verwenden.	Ich musste viele Dinge lernen, bevor ich mit dem System arbeiten konnte.	Ich musste eine Menge Dinge lernen, bevor ich mit diesem Produkt loslegen konnte.
	Strongly disagree – Strongly agree	Stimme <u>gar</u> nicht zu – Stimme voll zu	Stimme überhaupt nicht zu – Stimme voll zu	<Not Specified>	Stimme überhaupt nicht zu – Stimme voll und ganz zu

Gao et al. perform a reasonable validation and describe the translation process in general, but not in detail for every language. Therefore, it remains open if native speakers were involved in the translation of the German version and at which point Google Translate was involved. The translation is really close to the English version, which results in some strange word choices such as the translation of 'technical person' in item 4 or 'confident' in item 9. Overall, the wording seems unnatural. This seems to be the case for their French translation too: Gronier et al. translated the SUS into French and validated it comparing a word-for-word translation and a literary translation [2]. The word-for-word translation is similar to the French translation by Gao et al. but the literary translation is closer to a natural formulation in French. They found the literary translation to be more suitable.

Independent of the comparison results, all four translations have been applied in multiple studies. The translation by Rauer, for instance, is used by Berkemeier et al. [13], Küppers et al. [14] and is the SUS version presented by Wikipedia [15]. Tyers and Krantz started a project to translate the SUS into multiple other languages and validated each of them [16]. As German translation, they suggest the version presented by Rummel et al. However, until now, no validation was performed and the references are missing or are malfunctioning. Among others Bockhacker et al. [17] and Schubhan et al. [18] used this translation to evaluate their systems. The version of Lohmann and Schäffer is used in different projects to measure their users satisfaction or the usability of their system [19,20]. The German translation of Gao et al. was used for example to evaluate the usability of a data capture app [21] or to determine the feasibility of exergames in rehabilitation [22].

In conclusion, all three translations, except by Rauer, have their strength and weaknesses. Based on subjective preferences and individual weightings between natural language and creation process all four translations may be the favorite. If a peer-reviewed validation is considered mandatory, only the version of Gao et al. is valid alternative. Otherwise, although it is not fully validated, we suggest using the translation by Rummel et al. Its development process is methodically described the cleanest. The version of Lohmann and Schäffer is the close runner up, as it may improve the wording at the expense of methodological accuracy.

This work has limitations. German translations might exist that were not found during the search. In addition, it is still necessary to validate the proposed version as our analysis as non-language experts might not detect fine details of translation issues.

5. Conclusion

The search for German translations of the System Usability Scale yielded four translations created for further use. The comparison of the four versions revealed weak documentation of the translation process, missing validation for three of the versions and differences in the wording. The only validated version by Gao et al. seems to have a unnatural wording. The proposed version by Rummel et al., although not validated, is a good compromise between wording and methodically clean development.

References

- [1] Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability Evaluation In Industry*: CRC Press; 1996, p. 207–12.
- [2] Gronier G, Baudet A. Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale. *International Journal of Human-Computer Interaction* 2021;37(16):1571–82.
- [3] Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies* 2009;114–23.
- [4] Finstad K. The System Usability Scale and Non-Native English Speakers. *J. Usability Studies* 2006:185–8.
- [5] Gao M, Kortum P, Oswald FL. Multi-Language Toolkit for the System Usability Scale. *International Journal of Human-Computer Interaction* 2020;36(20):1883–901.
- [6] Dianat I, Ghanbari Z, AsghariJafarabadi M. Psychometric properties of the persian language version of the system usability scale. *Health promotion perspectives* 2014;4(1):82–9.
- [7] Berke L, Huenerfauth M, Patel K. Design and Psychometric Evaluation of American Sign Language Translations of Usability Questionnaires. *ACM Trans. Access. Comput.* 2019;12(2):1–43.
- [8] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)* 2021;372:n71.
- [9] Rummel B. System Usability Scale – jetzt auch auf Deutsch. [May 31, 2023]; Available from: <https://blogs.sap.com/2016/02/01/system-usability-scale-jetzt-auch-auf-deutsch/>.
- [10] Lohmann K, Schäffer J. System Usability Scale (SUS) – An Improved German Translation of the Questionnaire. [June 09, 2022]; Available from: <https://web.archive.org/web/20160330062546/http://minds.coremedia.com/2013/09/18/sus-scale-an-improved-german-translation-questionnaire/#6>.
- [11] Rauer M. Quantitative Usability-Analysen mit der System Usability Scale (SUS). [May 31, 2023]; Available from: <https://blog.seibert-media.net/blog/2011/04/11/usability-analysen-system-usability-scale-sus/>.
- [12] Reinhardt W. Crowdsourcing the translation of SUS. [May 31, 2023]; Available from: <https://web.archive.org/web/20160321223822/http://isitjustme.de/2012/01/crowdsourcing-the-translation-of-sus/>.
- [13] Berkemeier L, Werning S, Zobel B, Ickerott I, Thomas O. Der Kunde als Dienstleister: Akzeptanz und Gebrauchstauglichkeit von Smart Glasses im Self-Service. *HMD* 2017;54(5):781–94.
- [14] Küppers B, Schroeder U. System Usability Scale for FLEX: RWTH Aachen University; 2021.
- [15] Wikipedia. System Usability Scale. [May 31, 2023]; Available from: https://de.wikipedia.org/wiki/System_Usability_Scale.
- [16] Tyers B, Krantz P. SUS-translations. [May 31, 2023]; Available from: <https://github.com/ei8fdb/SUS-translations>.
- [17] Bockhacker M, Syrek H, Elstermann von Elster M, Schmitt S, Roehl H. Evaluating Usability of a Touchless Image Viewer in the Operating Room. *Applied clinical informatics* 2020;11(1):88–94.
- [18] Schubhan M, Altmeyer M, Buchheit D, Lessel P. Investigating User-Created Gamification in an Image Tagging Task. In: Bernhaupt R, Mueller F'F', Verweij D, Andres J, McGrenere J, Cockburn A, et al., editors. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM; 2020, p. 1–12.
- [19] Budurushi J, Renaud K, Volkamer M, Woide M. An investigation into the usability of electronic voting systems for complex elections. *Ann. Telecommun.* 2016;71(7-8):309–22.
- [20] Strickroth S. PLATON: Developing a Graphical Lesson Planning System for Prospective Teachers. *Education Sciences* 2019;9(4):254.
- [21] Reichold M, Heß M, Kolominsky-Rabas P, Gräßel E, Prokosch H-U. Usability Evaluation of an Offline Electronic Data Capture App in a Prospective Multicenter Dementia Registry (digiDEM Bayern): Mixed Method Study. *JMIR formative research* 2021;5(11):e31649.
- [22] Altorfer P, Adcock M, Bruin ED de, Graf F, Giannouli E. Feasibility of Cognitive-Motor Exergames in Geriatric Inpatient Rehabilitation: A Pilot Randomized Controlled Study. *Frontiers in aging neuroscience* 2021;13:739948.