System Usability Scale Formula as Alternative for Calculating Composite Score of Likert-type Items

Donald Stephen, Shahren Ahmad Zaidi Adruce, Ng Giap Weng
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak
Kota Samarahan, Malaysia
donaldstephen89@gmail.com, azshahren@fcs.unimas.my, gwng@fcs.unimas.my

Abstract— **Since its conception, Likert Scales have been used extensively in various fields. There are many ways to describe data obtained with Likert scales. In Human Factors research, System Usability Scale (SUS) is a widely adapted Likert Scale instrument to assess the usability of various products and systems. Researchers often have the tendency to modify SUS to fit their context of use, leading to various versions of SUS. Calculation of the composite score for SUS instrument does not involve computation of arithmetic mean/sum and re-coding of statements. SUS formula obtains the score for each Likert-type items and convert them into a single scale range from 0 to 100. Although SUS is a popular instrument, there is no formula introduced to facilitate the calculation of scale. This article aims to introduce a formula to facilitate the calculation of various versions of SUS with different number of items, number of positive and negative statements, as well as scale size. Although primarily meant for SUS practitioners, the formula presented can be used for any instrument adapting Likert-type items should the researchers decide to transform responses into 0-100 scale instead of computation of mean or sum for various statistical analyses. The controversial computation of mean for Likert Scales has been subjected to decades of debate. This article demonstrates an alternate method to calculate composite score of Likert-type items based on calculation of SUS. Results show the formula yield the same results and perfect relationship with mean and sum for various tests, indicating it can be used as an alternative to mean and sum as a composite score of Likert-types items.**

Composite Score; Likert Scales; Likert-type Items; System Usability Scales (SUS); Usability

I. INTRODUCTION

Rensis Likert (pronounced /ˈlɪkərt/) first introduced the idea of Likert scale in 1932 as a new technique for attitudinal research [1], [2]. Initially used for psychological and social researches, the use of Likert scales have extended to various fields, such as computational and medical sciences. Commonly, respondents are required to rate their level of agreement or disagreement with a given statement [3]. Researchers can also adapt different response anchors such as level of acceptability, importance, desirability, and frequency [4]. Likert scales give ordinal data as it is ranked from one extreme to another extreme [5]. Likert scale data is obtained by combining four or more Likert-type items measuring the same construct. This is done by calculating the composite score (mean or sum) of those Likert-type items [6].

Likert Scales has gained significant popularity since its conception and gave rise to various instruments. It has been adopted vastly over the years, thus became one of the most popular scaling methods because of its simplicity and robustness. One of the most notable uses of Likert Scales in Usability Studies is System Usability Scale (SUS). SUS is a popular, remarkably simple and reliable tool to measure the usability of diverse type of interfaces. SUS is a non-proprietary 10-items questionnaire developed by Brooke in 1996 to assess how people perceive the usability of a system [7], [8]. In SUS, respondents are required to rate their degree of agreement or disagreement with a set of statements on a 5 or 7 point scale [7]. Typically, the positive and negative statements in SUS are arranged alternately. The composite score of SUS is then calculated in a unique way, producing a single score which ranges from 0 to 100 [7]. Although principally developed for systems, SUS is flexible enough to be implemented for a wide range of interface technologies [9]. The use of SUS in usability testing has notably increased. It has been cited in more than 1,200 publications in 2013, not inclusive of unpublished usability tests [8].

The original SUS consists of both positive and negative statements and customarily arranged in alternate manner [10]. The calculation of standard SUS requires the researcher to identify negative and positive statements because they are treated differently in the calculation. The calculation leads to the overall value of usability which ranges from 0 to 100 [8]–[11]. The higher the SUS score, the higher is the perceived usability. The range of 0 to 100 obtained corresponds to what is called visual analog scale (VAS) whereby it defines exclusively the poles of a continuum (0 for lowest, 100 for highest) with no verbal label attached to each value. VAS can be regarded as equidistant and is on the level of an interval scale [12], [13]. Therefore, data corresponds with VAS can be analyzed using a wide range of statistical procedures [13].

II. RATIONALES

Researches were divided on how to treat Likert scales correctly. There are controversies in interpreting Likert scales, on whether they can be treated as interval data and subsequently used in parametric tests [3]. One of the major arguments on the treatment of Likert scales is the computation of mean and standard deviation to draw conclusions about the data. Likert scales have a rank order, and therefore should be treated as