Can percent agreement be the scoring scheme for safety attitudes questionnaire?

For more than a decade, the Safety Attitudes Questionnaire (SAQ) has been one of the most popular safety culture measurement instruments around the world. This short editorial addresses the scoring system that the original rubric of SAQ suggested. For a smoother explanation, we use a teamwork climate (TC) domain that consists of five items in SAQ. Each item is measured with a 5-point Likert scale (1=Disagree Strongly, 2=Disagree Slightly, 3=Neutral, 4=Agree Slightly, 5=Agree Strongly). These 1-5 raw scores are then converted into a 0-100 scale with 25 as the gap between the nearest scores. For each respondent, the average of the above five scores is calculated and, naturally, also ranges from 0-100.

With these individual participants' scores, we calculate a unit-level score—the hospital level is determined in the same manner. The original rubric of SAQ for TC describes the unit-level scoring as follows: First, calculate the individual respondent's mean TC domain score (0-100); second, dichotomize the mean scores equal to or above 75 and below 75. The former is called 'people agree teamwork is good,' and the latter is 'people disagree teamwork is good.' Finally, "Percentage Agreement" is obtained by dividing the number of 'people agree' by the number of total respondents multiplied by 100 to obtain a percentage scale.

This approach was particularly useful, especially when SAQ was first developed, in the era when safety was not a widespread topic of interest. At that time, the priority was getting more people to embrace the idea of safety and, thus, the percentage agreement served as the measure for such level of infiltration. Of course, computational convenience also contributes to some degree to its popularity. However, the salience of a safety culture including a teamwork climate has dramatically grown for the past few years, as has the computing environment. In the current realm of multidimensional item response theory, which allows an unprecedented level or granularity in the safety culture measurement, the above-mentioned simple scoring scheme may have lost its practical value.

Besides computing power, let us look into the innate problem of the original scheme with the following two examples. First, if a respondent assigns 4 (Agree Slightly) to all five items this year, her score is 75 and she must be treated as a person who agrees with the TC; thus, she is counted in the numerator of percentage agreement. On the other hand, if she assigns 4 to the first four items and 3 (Neutral) to the last item next year, her TC score will go down to 70; thus, she will be categorized as a person who disagrees with the TC. Is this reasonable? Can we really judge her as a person who agreed and disagreed with TC based on the difference in response to just one item among five between Agree Slightly and Neutral? Is it rational that she suddenly became a different person with a different perspective on TC? Maybe it is possible, but the logic is not persuasive. Maybe in year 2, she is not in a good mood, under the weather, or perhaps having a little trouble with her colleagues, leading her to give a slightly lower rating

Hsun-Hsiang Liao,[1] Wui-Chiang Lee,[2] Heon-Jae Jeong[3]
[1]Deputy Executive Officer, Joint Commission of Taiwan, Taiwan
[2]Department of Medical Affairs and Planning, Taipei Veterans General Hospital & National Yang-Ming University School of Medicine, Taipei, Taiwan
[3]The Care Quality Research Group, Chuncheon, Korea

Correspondence: Wui-Chiang Lee, Department of Medical Affairs and Planning, Taipei Veterans General Hospital & National Yang-Ming University School of Medicine, Taipei, Taiwan, Tel +886-2-28757120, Fax +886-2-28757200, Email leewuichiang@gmail.com

for just one item. Due to this excessive dichotomization, the original SAQ scoring scheme gives her the title of a 'person who does not agree with the TC.'

The trickier part is that such change in a person's agreement status can be double counted: In the above example, the number of people agreeing with TC decreased by one (her), which is completely fine. However, people who disagree with TC increased by one (her) simultaneously. If you use proportion-based statistics, this might be okay, but if we rely on our favorite odds ratio (OR), this is certainly a double dip. In other words, the change in OR is inflated.

Now, we move on to another issue, namely, distribution: Instead of individual respondents, let us assume we are dealing with a clinical arena with 100 healthcare professionals. The score distribution happens to be bimodal, which means one hump in the high-score region, another in the low-score region.[1] Then regardless of what happens below the score of 75, the percentage agreement is the same. Even though the scores of people in lower levels really improved to somewhere around the middle, which is a phenomenal change in a good way, the change cannot be captured with this traditional approach. The other way around is also possible.

We have been helped by SAQ and its original scoring scheme. There is no doubt they changed healthcare, making it safer. However, as with all medical science, if new methodology develops, the previous one should honorably retire. As of now, the scoring methods for safety culture measurement instruments demonstrate this very case. We have to develop a new scoring methodology for SAQ and other instruments that can provide information about both centrality and spread with exceptional granularity. Luckily, we have several potential arsenals, some of which came from other fields such as psychology and education. We can begin with trying them, and

based on our preliminary studies, the potential methods are quite promising.[2-4] We do not have to re-invent the wheel.

Many of us might think, "The old method is good enough, although a more precise one has come out." This is a common reaction from all of us; we all are hesitant to make changes, even if they are supported by solid evidence. However, the field we work in castrates this hesitation. Let's ask ourselves, "If we switch the scoring scheme, we can save many more lives. Can we handle the guilty conscience from not switching it?" We know the answer, the same answer.

Of course, we will provide step-by-step guidelines to achieve new scores without adding any work. Until then, stay tuned. The wait will not be long.

Acknowledgements

Conflicts of interest

None.

References

1. Jeong HJ, Lee WC. A Very Short and Gentle Review of Mean and Standard Deviation as Summary Statistics of a Sample. *Biometrics & Biostatistics International Journal*. 2017;6(2):00160.

2. Jeong HJ, Jung SM, An EA, et al. Combinational Effects of Clinical Area and Healthcare Workers' Job Type on the Safety Culture in Hospitals. *Biom Biostat Int J*. 2014;2(2):00024.

3. Jeong HJ, Lee WC. The Pure and the Overarching: An Application of Bifactor Model to Safety Attitudes Questionnaire. *Biom Biostat Int J*. 2016;4(6):00110.

4. Jeong HJ, Lee WC. Item Response Theory-Based Evaluation of Psychometric Properties of the Safety Attitudes Questionnaire—Korean Version (SAQ-K). *Biom Biostat Int J*. 2016;3(5):00079.