Available online at www.sciencedirect.com

ScienceDirect

Currents in Pharmacy Teaching and Learning 7 (2015) 836–850

Currents in Pharmacy Teaching & Learning

http://www.pharmacyteaching.com

Review article

How to analyze Likert and other rating scale data

Spencer E. Harpe, PharmD, PhD, MPH*

Midwestern University Chicago College of Pharmacy, Downers Grove, IL

Abstract

Rating scales and rubrics are commonly used measurement tools in educational contexts. Unfortunately, there is a great deal of controversy surrounding how data derived from these tools can and should be analyzed. One issue that is repeatedly raised is whether these data are ordinal or continuous. A related question is whether parametric data analysis techniques are appropriate and/or acceptable for these rating scale data. Some of this controversy may stem from a misunderstanding of fundamental issues related to these particular tools or a poor use of terminology. This article provides a review of basic issues surrounding measurement of various phenomena relevant to educational settings, as well as previous empirical studies examining the effects of using parametric analysis approaches on rating scale data. Based on previous empirical evidence reviewed in this article, parametric analytical approaches are acceptable provided certain criteria are met. Implications for research and teaching are also briefly discussed. After reading this article, the reader should be able to identify the characteristics of a true Likert scale and explain the situations when parametric analytical techniques are potentially appropriate for rating scale data or when nonparametric techniques are preferred.

Keywords: Likert scales; Data analysis; Measurement; Summated scales; Rating scales; Ordinal data

Situation

When I took my first academic position and became involved in educational scholarship, I found myself asked to justify some of my analytical decisions. This was not a bad thing in and of itself since we must be able to justify our choices in any scholarly endeavor. At the time, what struck me as odd were the comments related to how I had analyzed data from rating scales. I remember one passerby at a poster session stopping to look at my poster and saying, "That is interesting, but you know you really should not have used means to report Likert scale results. After all, those are ordinal data." I was slightly taken aback but thanked the individual for his/her comments and started to think. The statement about Likert scales being ordinal data was never made in my statistics classes in graduate school, so I wrote the comment off as a one-time occurrence. Suffice it to say that was not a one-time occurrence. As a peer reviewer for journals, I sometimes see comments from other reviewers along the same lines as my poster commenter. This issue has also arisen when working with pharmacy practice residents developing their residency projects. After a brief look through the literature, it was clear that the issue of how to analyze data from rating scales and rubrics is not altogether straightforward. The purpose of this article is to provide a review measurement as it relates to the educational context and to provide some evidence-based recommendations for analytical approaches.

Although this article focuses on recommendations for analyzing data from various rating scales for self-reported concepts like self-confidence with a skill learned in class or student perceptions of learning, the discussion herein applies equally to instruments developed to guide external assessment or evaluation (e.g., preceptor evaluation of student performance on an advanced pharmacy practice experience (APPE) or a peer-evaluation of an instructor's lecture). Nothing in this article is meant to contradict the

* Correspondence to: Spencer E. Harpe, PharmD, PhD, MPH, Chicago College of Pharmacy, Midwestern University, 555 31st Street, Downers Grove, IL 60515.

E-mail: sharpe@midwestern.edu

http://dx.doi.org/10.1016/j.cptl.2015.08.001

1877-1297/

recommendations Peeters provides for constructing rubrics to capture rater judgments.[1,2] While that article focuses on the construction of rubrics, the recommendations from this article could be used for the analysis of those ratings.

One word of caution is needed. This article focus on measurement and analysis in the quantitative paradigm that should not be misconstrued as a dismissal of the importance or utility of qualitative research methods. To the contrary, qualitative methods are extremely important and can provide a depth of information that quantitative measurement cannot begin to approach. Information gathered from qualitative methods can inform quantitative methods, and vice versa. Furthermore, both methods can be successfully used together as seen in mixed-methods studies.[3]

Methodological literature review

Measurement forms an important part of research and evaluation efforts in the quantitative paradigm insofar as we are concerned with the magnitude, or at least the relative magnitude, of some phenomenon of interest. It is no surprise that the process of measurement is relatively straightforward for physiological quantities (e.g., blood pressure or cholesterol) or for certain clinical outcomes (e.g., hospital length of stay or number of on-time prescription refills). The challenge that we face in education is that many of the phenomena we seek to measure are not physical but are cognitive in nature. The question becomes how can we measure these non-physical phenomena. Some may be directly observable and readily quantifiable, such as the number of times a student commits an error when counseling a patient or the number of seconds an instructor waits for student responses after posing a question during lecture. Other phenomena may be observable yet not directly quantifiable, per se. For example, a preceptor can provide a qualitative assessment of a student's performance of some skill or activity (e.g., "excellent" or "below average"), but a number is not directly observable. Finally, other phenomena, such as self-confidence, are not directly observable at all and must be assessed by self-report. A wide variety of phenomena that are of particular interest in education fall into these latter two categories. Numerous rating scales and rubrics have been developed to allow us to derive quantitative measures of non-physical phenomena by combining a set of items asking an individual to make a series of qualitative assessments. Despite the availability of these instruments, there has been considerable confusion when analyzing data derived from them. The perceived misuse of certain forms of statistical analysis with these instruments has been lamented in the literature,[4,5] yet the recommendations for analysis have stirred considerable controversy.[6–11] The following sections provide a review of the historical development of measurement and associated issues with statistical analysis.

A brief history of measurement

Dating back to the ancient Greeks, measurement is not a recent concept. The modern view of measurement is generally built on Cartesian thinking and was developed during the scientific revolutions of the 15th to 17th centuries. From Descartes' philosophical stance, all *physical* properties could be represented by some quantity. The implication was that quantification was the only option for the scientific method, at least for the natural and physical sciences, thus a quantitative imperative came to be promoted in order for an endeavor to be considered "scientific."[12] This imperative can best be seen when William Thomson (aka, Lord Kelvin) stated that "when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of an unsatisfactory kind."[13]

The effects of this quantitative imperative on psychological and sociological sciences, and subsequently on educational science, carried into the early 20th century. In the 1930s, the British Association for the Advancement of Science summoned a committee to examine the status of psychophysical measurement methods.[12] In a manner that could be compared to the way a child's performance in school is compared to an older sibling, psychological measurement methods were compared to measurement methods in "older" sciences like physics. The committee generally used Campbell's theory of measurement. This theory described two types of measurements: fundamental measurements and derived measurements.[14] Since derived measurements depended on (i.e., were derived from) fundamental measurements, Campbell viewed fundamental measurements as a necessary step for quantification.[12] Using Campbell's theory, the British committee concluded that psychophysics did not have any "true" measurement methods because there were no fundamental measurement methods. Without a measurement method, psychophysics, and by extension all of psychology, could not be considered a science.[12,15] Simultaneous with the committee's work, and possibly in anticipation of their final report, Stevens set forth a framework for psychological measurement. With this development, psychology could no longer be classified as a non-science since there was now a system of measurement. The implications of Stevens' measurement framework measurement are far-reaching since he proposed the four levels, or scales, of measurement that we continue to use today—nominal, ordinal, interval, and ratio.[16]

In the early-to-mid 20th century, researchers developed various scaling methods to measure the relative strength or degree of psychological phenomena, such as mental abilities and attitudes.[17] As Stevens put rather simply, scaling was a process to assign a number to objects using some rule.[16] Through defining measurement scales, any phenomenon could be "quantified." The development of methods to measure attitudes resulted in several noteworthy scaling