Chapter X
Equivalence of Electronic and Off-Line Measures

Lynne D. Roberts
Crime Research Centre, University of Western Australia, Australia

ABSTRACT

This chapter explores the measurement equivalence of paper-and-pencil and electronic surveys and measures. After reviewing competing claims and the limitations of the research, there is a strong case that many measures do retain their psychometric equivalence across modes of administration, with the exception of speeded tests and measures that may be subject to social desirability response sets.

INTRODUCTION

Paper-and-pencil surveys and measures are frequently transformed into electronic formats for use. Many assume that because the questions remain identical that the two modes of administration can be used interchangeably, but is this the case? Does the mode of administration of surveys and measures make a difference to the results obtained? This chapter explores the measurement equivalence of paper-and-pencil and electronic surveys and measures. First, measurement equivalence is defined and methods of determining equivalence outlined. Next is an examination of the research literature pertaining to the extent and type of differences that have been found by modes of administration, and the reasons hypothesized to account for these differences. This is followed by an examination of the limitations of the research conducted to date. Finally, recommendations are provided for the electronic administration of paper-and-pencil surveys and measures.

BACKGROUND

Historically, many measures and surveys have been created in a paper-and-pencil format for interviewer or self-administration. The introduction of computers to the academic community enabled attempts to computerize many measures (Epstein & Klinkenberg, 2001). With the introduction of the Internet, a further stage in the evolution of measures has developed: the electronic distribution of surveys and measures. The introduction of Hypertext Markup Language 2 (HTML 2)

What is Measurement Equivalence?

Measurement equivalence refers to the stability of a measure’s factor structure across situations. When paper-and-pencil measures are converted for use in electronic environments, it is essential to maintain the psychometric properties of the measure. In order to determine measurement equivalence across modes of administration, an assessment of means, standard deviations, and structures resulting from the two modes of administration are required. In addition, the two results should be highly correlated to ensure they are both measuring the same concept (Taris, Bok, & Meijer, 1998). Statistical methods used to study the structural invariance of measures are confirmatory factor analysis and item response theory (Meade & Lautenschlager, 2004; Raju, Laffitte, & Byrne, 2002). A detailed account of the procedures involved in establishing measurement equivalence is beyond the scope of this article, but see Taris and colleagues (1998) for an eight-step procedure for determining structural invariance.

Are Electronic and Paper-and-Pencil Measures Equivalent?

Early measurement equivalence research compared paper-and-pencil and pre-Internet computer administered measures and surveys. The results obtained varied by the type of measure. Mead and Drasgow (1993) conducted a meta-analysis of computer vs. paper-and-pencil cognitive ability tests, based on 29 studies and 159 correlations. They concluded that speeded tests were affected by mode of administration (moderate effect size), but that timed power tests (tests that assess ability within a practical limitation of time) were not. A suggested reason for the differences in speeded tests by mode of administration was the different motor skills required for computer and paper-and-pencil tests. In contrast to the cognitive tests, no differences were found in work-related noncognitive tests (King & Miles, 1995) and only minor differences found in nonaptitude psychological measures (Rosenfeld, Booth-Kewley, Edwards, & Thomas, 1993). Most differences found in personality tests such as the MMPI, EPI, 16PF, and CPI were attributable to the different response options offered, rather than the actual mode of administration, highlighting the need for instructions, presentation, and response formats to be kept equivalent (Rosenfeld et al., 1993). The measurement equivalence of computerized adaptive testing (where the level of subsequent items is determined by previous responses) has yet to be established (Epstein & Klinkenberg, 2001).

Survey responses may also be affected by mode of administration. Liefeld (1988) found that similar responses were obtained on multiple-choice and yes/no questions in computer-administered, personal interview, and self-completion measures. However, for questions where multiple responses could be selected, higher numbers of responses were selected in the computer-administered condition. Similarly, Mehta and Sivadas (1995) reported no differences in classification or attitudinal variables by mode of administration (e-mail vs. paper-and-pencil), but found that e-mail respondents write more comments than paper-and-pencil respondents. Keyboard design, screen size, screen illumination, and processing speed of respondents’ computers do not affect survey responses (Rosenfeld, Doherty, Vicino, Kantor, & Greaves, 1989). However, responses can be affected by the programming of measures, such as the ability to scan, preview, review, or change answers (Kiesler & Sproull, 1986; Whitener & Klein, 1995). Thus, the programming of surveys and measures appears of more importance than the actual computer hardware used by research participants.