Reliability is "Trivial": A Cautionary Tale of Interpreting Reliability of Sport Management Scales

Luke Mao, University of New Mexico
Min Jung Kim, University of New Mexico
Yen Hsun Chen, University of New Mexico
Yong Wang, University of New Mexico

Method - Quantitative (Other) Saturday, June 1, 2019
20-minute oral presentation (including questions) 3:10 PM
Abstract 2019-418 Room: Oakley

The growth of sport management discipline has been accompanied by an accumulated literature on measuring sport management specific constructs. A review of existing scale development practices in sport management revealed that the true-score based classical testing theory and factor analytic model remain the predominant paradigm to develop scales and the concepts of reliability and validity are integral to the scale development process. All researchers have discussed about measurement properties in the method section and the Cronbach's alpha probably is the most circulated statistic in the field. Upon further review of the reported alpha coefficient scores, very few researchers seemed to have failed the reliability test by demonstrating that alpha reached to the threshold of 0.7. Most published scales actually have achieved high alpha scores. Reliability seems to be a rather "trivial" empirical task!

The triviality of reliability test is a puzzle. On the one hand, experiences inform us that studies are difficult to replicate (i.e., low reliability); on the other hand, almost all studies have reported acceptable reliability. Furthermore, reliability estimates, including parallel-test, test-retest, Cronbach's alpha, are the lower bound estimates of reliability (Sijtsma & van der Ark, 2015), which suggests that the actual unobservable population reliability would be higher. However, Cronbach's alpha as an estimate of reliability is conditioned on the assumptions of: 1) classical true score; 2) tau-equivalency of items (i.e., equal loading on an underlying factor); and 3) uncorrelated errors (Novick & Lewis, 1967; Yang & Green, 2011). Assumption 1 is philosophical; a violation of assumption 2 makes estimates even more conservative; and only a violation of assumption 3 makes estimates spuriously higher (Green & Yang, 2009). Several sources contribute to correlated errors, one of which is transient errors due to unstable feelings, attitudes and mental awareness of the respondents (Becker, 2000).

To demonstrate the puzzle, reliability of 18 sport management scales was evaluated by the test-retest method. Reliability estimate based on this method has its own weakness and is usually 0.1 smaller than the Cronbach's alpha. However, it is a common misunderstanding that the two methods estimate two different aspects of reliability, i.e., internal consistency v.s. stability; they are estimates of the same reliability. Test-retest has the advantage of examining the transient error problem. A total of 126 participants rated 113 items selected from 18 scales repeatedly with an interval of two weeks. Intraclass correlation coefficients (ICC) was used to measure reliability (Weir, 2005). The results suggested there is a significant gap between test-retest and Cronbach's alpha estimates. The item reliability from Cronbach's alpha ranged from .53 to .98 (M=.83, SD=.12); construct reliability ranged from .74 to .98 (M=.83, SD=.07). By contrast, the item reliability from test-retest ranged from .09 to .92 (M=.60, SD=.17); construct reliability ranged from .25 to .84 (M=.67, SD=.13). The discrepancy between these two estimates is large and alludes to some neglected areas in scale development: sources of correlated errors. The apparent triviality of reliability is nontrivial: cautions should be exercised when interpreting reliability estimates and more rigorous procedures should be taken to develop a scale.