Scale Development Using Test-retest Reliability: Implications for Consumer Research

Yiqi Yang, University of Ottawa
Marijke Taks (Advisor), University of Ottawa
Eric MacIntosh (Advisor), University of Ottawa

Method - Quantitative (Other) virtual asynchronous
20-minute oral presentation (including questions) Session: Method/Consumer Behavior

Marketing research seeks to collect and produce reliable data to provide decision makers with the best information possible to make predictions, which is essential for marketers and quality data interpretation. Although some errors will be present in almost any measurement processes, reliability tests play a critical role in determining data quality. The difficulty is that methods of measuring reliability vary and there is no guarantee that the same results will occur when different methods are applied. Although internal consistency reliability is used more frequently than test-retest or alternate-form techniques in sport marketing research designs, test-retest has the advantage of examining the transient error problem, and there is a significant gap between test-retest and Cronbach's alpha estimates (Mao et al., 2019). Previous studies on ski participation failed to report reliability scores. Therefore, the purpose of this study is to use both, test-retest and internal consistency methods to develop a more reliable instrument to measure the ski participation of a Canadian sample.

Based on Shank’s and Lyberger’s (2015) sports consumption model, an online questionnaire was developed, consisting of internal (21 items), external (18 items) and situational factors (15 items). Test-retest reliability was examined by having the same participants completing the questionnaire twice with a time interval of approximately two weeks (Straub, Boudreau & Gefen, 2004). Participants were recruited from a university sample of students, staff and faculty. Ethics to conduct the study was granted. During the first data collection (T1), 155 participants filled out the questionnaire; 73 participants filled out the questionnaire a second time (T2). Pearson correlations and intraclass correlation coefficients were calculated to measure test-rest reliability (Chung et al., 1998). Of the initial 54 items, only 29 showed sufficient reliability (correlations ranging from .71 to .78). Items of internal and external factors showed higher correlation scores than items of situational factors. Subsequently, a Principal Component Analysis was performed on the 29 retained items leading to 5 factors solution. Internal consistency among the variables for each factor was examined; Cronbach’s alpha scores ranged from .75 to .91.

The low test-retest reliability scores for items of the situational factor are not surprising, given that situations vary by default (Shank & Lyberger, 2015), and may therefore be difficult to capture when measured in different moments in time. Thus, when testing situational factors, the use of internal consistency measures that include a minimum of three variables to measure a similar idea or construct, such as the alternate form or split-half method is most appropriate (Thomas et al., 2005). Similar to other research, the reliability of situational factors related to ski participation may be questioned (e.g., Facullant et al., 2008; Konu et al., 2011). This study illustrated how test-retest reliability significantly reduced the number of variables for internal and external factors. The presentation will discuss the merits and challenges in using the test-retest design.