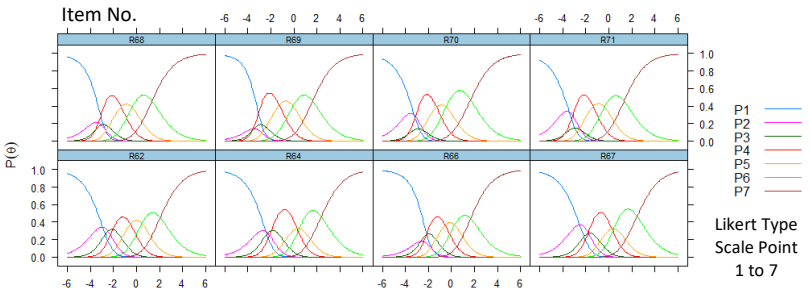
Measuring Human-Robot Interaction (HRI) in Fashion Stores: Scale Development Using Item Response Theory (IRT)

Christina Soyoung Song, Illinois State University, USA
Youn-Kyung Kim, University of Tennessee, Knoxville, USA
Bruce W. Jo, Tennessee Tech University, USA

Keywords: Fashion, Human-Robot Interaction**,** Item Response Theory

Background: The principles of Item Response Theory (IRT) deviate from Classical Test Theory (CTT), which has long been dominant in the field of scale measurement in fashion retailing. According to Embretson and Reise (2013), IRT posits that the probability of a participant providing a certain response to an item is influenced by both the participant's characteristics and the item's properties. Given the shortcomings of the CTT, which include the sample-dependent and central-tendency biased nature of derived scores and problematic model estimations in the presence of missing data, IRT has emerged as a preeminent psychometric approach for constructing and validating scales across diverse research domains, psychology, personality, health science, and large-scale educational assessments (Lord, 1980). Nevertheless, this fundamental concept of IRT statistical methods has been underused and has not been systematically introduced in fashion academia. To fill this gap and to encourage fashion scholars to explore the untapped potential of integrating IRT in scale development, this study first outlined the key tenets of IRT and Partial Credit Models (PCMs) of Rash IRT derived for multiple-choice (e.g., rating or Likert type scales). Second, through a solid literature review of the IRT-relevant studies, we demonstrated the necessity of bringing novel methodologies into our domain. Third, we used traditional CTT and IRT methods simultaneously in developing scale measurement. Fourth, we conducted personal interviews as a preliminary step to generate initial descriptions and perceptions of interacting with a service robot and extracted a definition of HRI as “the psychology of how humans interact with AI robots to accomplish collaborative tasks in stores.” Fifth, we analyzed data from 7-point Likert-type scale of the attitudes toward Human-Robot Interaction (HRI) that was developed as a 10-item scale consisting of emotional and cognitive HRI. Sixth, we provided a comparable interpretation through IRT/CTT and Item Characteristic Curves (Figure 1). Ultimately, this study evaluated the item quality of HRI and showcased the benefits of using CTT's Confirmatory Factor Analysis (CFA) and Rash IRT’s PCM simultaneously for creating and assessing new scale measurements containing polytomous items. This study offers evidence-based suggestions for integrating polytomous IRT and CTT's CFA



approaches to create a comprehensive and high-quality measurement scale, serving as an introductory primer.

Methods and Analytic Strategies: The study proposed the following methodological approaches and guidelines. The HRI instrument was conceptualized and designed to measure consumers' attitudes toward interaction with a service robot in fashion retail stores and consisted of emotional and cognitive HRI constructs. First, using a video-based stimulus, we initially developed 10 scale items of the attitudes toward HRI based on the studies of Nomura, Kanda, Suzuki, and Kato (2008) and Ko, Cho, and Roberts (2005) and conducted 14 personal interviews regarding the interaction with the service robots in fashion stores. The emotional HRI had 6 items, and the cognitive HRI had 4 items. Second, four faculty members at a major university in the Southeastern United States conducted a content analysis of survey items and the video-based stimulus. Based on their comments, the scale items and the stimulus were revised for clarity and readability. All scale items were measured on a 7-point Likert-type scale. Third, the Likert-type item evaluation using r v. 4.2.3 included Item Characteristic Curves (Figure 1), descriptive analysis (Figure 2), internal consistency, conversant validity, the combination of CTT'S CFA and PCM, a type of Rash IRT analysis for polytomous items (e.g., 7-point Likert scale) to assess the psychometric properties of the 10 HRI scale items (Figure 3). The two scale items from emotional HRI were excluded due to the distant meaning of emotion applied in the scale items and based on the results of an Exploratory Factor Analysis (EFA) and a CFA. Fourth, we conducted individual exploratory graph analyses (EGAs) for each emotional and cognitive HRI construct (Figure 4). Then, an EGA included the HRI constructs and core target variables of the intention to use a service robot and visit the robot-operated fashion stores (Figure 3). Finally, we analyzed Differential item functioning (DIF) to detect any DIF items with a different probability of responding to an item between genders.

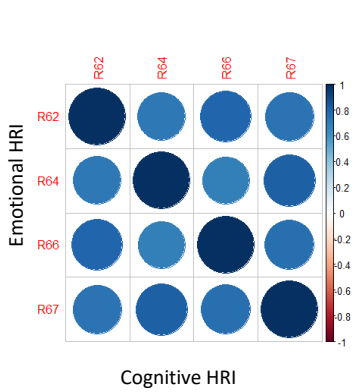


Figure 2. Correlation Heatmap

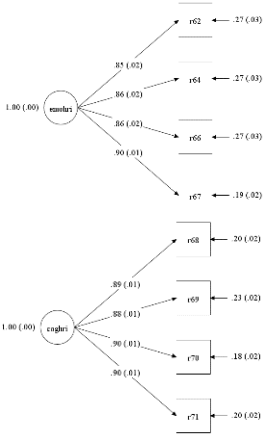


Figure 3. CTT CFA

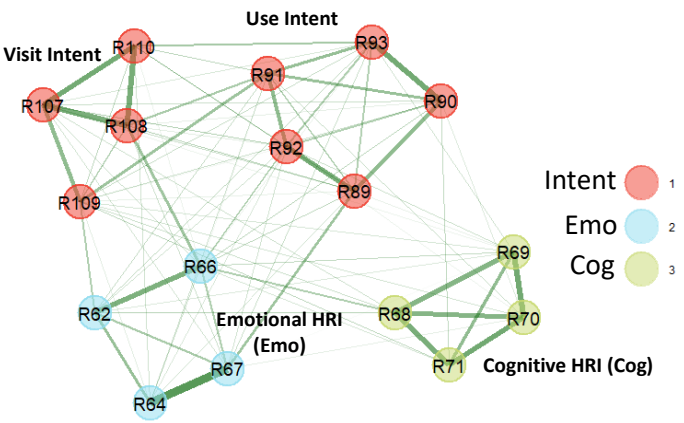


Figure 4. EGA Graph

Results: Using r v. 4.2.3, we conducted a series of CTT's EFA and CFA, Rash IRT's PCM, and EGAs to evaluate the scale measurement model of (1) emotional HRI and (2) cognitive HRI in fashion stores. The results of EFA indicated two underlying structures of observed variables of emotional and cognitive HRI. All CFA factor loadings were greater than .50 (range from .81 to .90; CFI = 0.943, GFI = 0.943), meeting the unidimensional assumption of IRT. The results of PCM IRT demonstrated excellent values of the item discrimination parameter (represented as *a*), ranging from 0.79 to 0.85. Further, no gender-based DIF was detected by IRT (Figure 1) Finally, EGA graph illustrates the appropriate item clusters and relations between the attitudes toward HRI and intentions to use the robots or visit the robot-operated fashion stores.

Conclusion: Through the utilization of an integrated approach to scale development, merging the IRT and CTT methodologies, this study provides substantial evidence that the employment of IRT and CTT's CFA offers distinctive insights into item quality, model fit, and scale performance. This methodological synthesis ultimately facilitates the production of precise and valid research instrumentation, thereby ensuring the generation of accurate findings. This study serves as a foundational guide to initiate an alternative process of developing a scale measurement.

References

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ko, H., Cho, C.-H., & Roberts, M. S. (2005). Internet uses and gratifications: A structural equation model of interactive advertising. *Journal of Advertising*, *34*(2), 57-70.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human--robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics*, *24*(2), 442-451.