The present study investigates psychometric properties of the items in Wann and Branscombe’s (1993) Sport Spectator Identification Scale (SSIS), which consists of seven items on an eight-point semantic differential scale. The SSIS has been widely used in sport management research for measuring the degree to which sport spectators feel psychologically connected to a team (e.g., James et al., 2019). Although several researchers examined the psychometric properties of the SSIS (e.g., Theodorakis et al., 2016), most of them focused on the usefulness of the SSIS at the scale level, but not at the item level. To address this gap, the present study examines the degree to which each item on the SSIS measures individuals’ team identification with statistical precision, using item response theory (IRT). The following research questions were addressed:

To what extent does each item on the SSIS reliably measure individuals’ team identification? If any items with low reliability are identified, what would be possible sources of error?

IRT is a modeling-based measurement theory that provides information about psychometric properties of items on a scale (Embretson & Reise, 2000). The fundamental idea of IRT is that an individual’s response to an item in a measurement instrument (e.g., survey) is a function of both the individual's underlying trait level and properties of the item. Employing IRT allows researchers to identify which items in a scale measure the underlying trait with precision and at which trait level each item has the maximum level of reliability.

A total of 635 spectators of three U.S. professional sports (i.e., the NBA, MLB, and MLS) were recruited via Amazon Mechanical Turk. Participants were asked to think about the most recent sporting event among the three leagues they attended over the previous month, and refer to one of the teams as the focal team when responding to the SSIS items.

The IRT analysis using the general partial credit model revealed that four items (Items 1, 2, 3, and 5) measured individuals’ levels of team identification with great precision, while three items (Items 4, 6, and 7) measured individuals’ levels of team identification poorly. Furthermore, in all seven items on the SSIS, response options 1 to 4 did not work as distinct options. Among eight response options within each item, we found that option 2 and option 4 were rarely the most likely option to be chosen across the trait levels.

Based on a close examination of the wording of Items 4, 6, and 7, we suggest that the low reliability of these items may be attributed to double-barreled questions (Items 4 and 7) and a negatively worded statement (Item 6). We made additional suggestions to address the low reliability issues with these items. We also suggest that the SSIS be revised to provide fewer than eight options within each item. Our findings and suggestions are expected to improve the measurement precision of the SSIS for both researchers and practitioners, and contribute to our understanding of the construct of team identification using the SSIS.