

9-2017

# Development of the Referee Retention Scale

Lynn L. Ridinger

Kyungun R. Kim

Stacy Warner

Jacob K. Tingle

Trinity University, [jtingle@trinity.edu](mailto:jtingle@trinity.edu)

Follow this and additional works at: [https://digitalcommons.trinity.edu/busadmin\\_faculty](https://digitalcommons.trinity.edu/busadmin_faculty)

Part of the [Business Administration, Management, and Operations Commons](#)

---

## Repository Citation

Ridinger, L. L., Kim, K. R., Warner, S., & Tingle, J. K. (2017). Development of the referee retention scale. *Journal of Sport Management*, 31(5), 514-527. doi:10.1123/jsm.2017-0065

This Post-Print is brought to you for free and open access by the School of Business at Digital Commons @ Trinity. It has been accepted for inclusion in School of Business Faculty Research by an authorized administrator of Digital Commons @ Trinity. For more information, please contact [jcostanz@trinity.edu](mailto:jcostanz@trinity.edu).

**Development of the Referee Retention Scale**

Re-Submitted on

May 7, 2017

### **Abstract**

Building upon the current sport officiating research, this study puts forth the Referee Retention Scale (RRS). Through a three-phase process, the researchers developed a valid and reliable scale to predict sport officials' job satisfaction and intention to continue. The first phase consisted of instrument development, while the second phase included field testing of referees (n=253). After EFA and Rasch analysis, the resultant refined scale from phase 1 and 2 was then administered to 979 referees in phase 3. Phase 3 results using CFA indicated that the 7-factor, 28-item RRS was a valid and reliable tool for measuring and predicting referee retention. The results highlight the importance of considering a variety of factors associated with the referee experience, which include Administrator Consideration, Intrinsic Motives, Mentoring, Remuneration, Sense of Community, Lack of Stress, and Continuing Education. A discussion on how the RRS can help administrators manage and retain sport officials is included.

**Keywords:** Referee, Scale Development, Job Satisfaction, Sports Management, Umpire

### **Development of the Referee Retention Scale**

Referees are vital to almost all organized sport competitions; however, sport managers and administrators are facing a growing problem as the number of qualified sport officials continues to decline (American Sport Education Program, 2011; Cuskelly & Hoye, 2013; Kim, 2016). A shortage of officials can negatively affect the quantity and quality of games being played. When officials are not available it is common for games to be cancelled or rescheduled (Topp, 2001). In the United States, some high school state associations have discussed reducing the number of games or dropping sports altogether because of a lack of officials (“Infographic,” 2016; Stevens, 2016). Moreover, the quality of the sporting experience for both players and spectators is negatively impacted when veteran officials are overworked (Cuskelly & Hoye, 2004), and novice officials are forced into situations beyond their current knowledge and skill level (Read, 2000). According to Barry Mano, the president of the National Association of Sports Officials, recruiting new referees continues to be difficult and there is a growing need to find ways to attract more men and women into officiating (Stevens, 2016).

The successful development of any organization or industry is dependent upon the recruitment, retention, and development of key personnel (Barr & Hums, 2012; Chelladurai, 2014). In the sport industry, referees play a key role in the workforce and it is important to understand factors associated with their retention so that sport organizers can develop strategies to more effectively recruit and retain referees. There is a growing body of literature related to various issues associated with sports officials. Such research includes patterns of initial entry into officiating (Furst, 1989, 1991; Purdy & Snyder, 1985); understanding stress, burnout, and coping responses of officials (Anshel, Kang, & Jubenville, 2013; Kellett & Shilbury, 2007; Rainey, 1995, 1999; Rainey & Hardy, 1999; Taylor, Daniel, Leith, & Burke, 1990); referee self-

efficacy (Myers, Feltz, Guillen, & Dithurbide, 2012); involvement with officiating (Ridinger, 2015); referee's sense of community (Kellett & Warner, 2011); organizational support (Cuskelly & Hoye, 2013; Kim, 2016); gender equity (Kim & Hong, 2016; Nordstrom, Warner, & Barnes, 2016; Schaeperkoetter, 2016; Tingle, Warner & Sartore-Baldwin, 2014); and attrition of sports officials (Forbes & Livingston, 2013; Warner, Tingle, & Kellett, 2013).

With the growing research on referees, a strong foundational knowledge base has been built. However, much of this work has been qualitative. From a practical standpoint, the industry would benefit from a tool that would help predict whether a sports official is likely to continue in the role, and measure the relevant factors that influence this decision. Building upon the extant sport officiating literature, there is a clear need to develop a valid and reliable instrument to measure variables that can predict referee retention. A few referee specific scales do exist. The Soccer Officials Stress Survey (Taylor & Daniel, 1987) and the Basketball Officials Sources of Stress Inventory (Kaissidis & Anshel, 1993) were precursors to Anshel et al.'s (2013) Sources of Acute Stress Scale for Sports Officials. Additionally, Myers et al. (2012) developed the Referee Self-Efficacy Scale. Each of these scales focused on a specific psychological variable, yet recent research has shown that multiple sociological and organizational factors might combine to have a greater impact on referees' decisions to either stay or discontinue officiating (Cuskelly & Hoye, 2013; Forbes & Livingston, 2013; Kim, 2016; Kellett & Warner, 2011; Ridinger, 2015; Tingle et al., 2014; Warner et al., 2013).

As a result, the existing referee scales that have focused on stress (i.e., Anshel et al., 2013; Kaissidis & Anshel, 1993; Taylor & Daniel, 1987) or self-efficacy (Myers et al., 2012) fall short of capturing the entire referee experience. In fact, Kellett and Shilbury's (2007) work highlighted that the anecdotally reported stress and abuse did not seem to contribute to referee attrition, while

Phillips and Fairley's (2014) work positioned officiating as a serious leisure pursuit rather than a stress inducing job. The sum of this recent work points to the multifaceted nature of sports officiating and consequently, a need for a multidimensional tool to predict referee retention. Given the current state of knowledge regarding refereeing, it is clear that this tool needs to go beyond just focusing on the psychological variables and better encompass all stages of the referee experience. Therefore, this study sought to develop a valid and reliable multifaceted scale to measure factors that have been reported to affect referee retention and attrition.

One of the more comprehensive studies on officiating was Warner et al.'s (2013) work with former basketball officials. That study was significant because it examined the entire referee experience from initial entry through discontinuation from the role. The purpose of their research was to explore the experiences of former referees with the aim of identifying strategies that would have ultimately led to their retention rather than their departure from the role. Warner and colleagues used a phenomenological approach to discover and critique the experiences expressed by 15 former referees regarding their involvement from start to finish with the officiating profession. They utilized Green's (2005) Sport Development Model as a framework to examine three stages of referee development - Referee Recruitment, Referee Retention, and Referee Advancement. By taking such a sport-focused approach, the distinctiveness of the role of sport officiating was seemingly captured (*cf.* Chalip, 2006). Because the Warner et al. study focused on those who had actually left the refereeing profession, aspects that impacted decisions to leave were highlighted and the resulting conceptual framework was called the Referee Attrition Model. This model included dimensions influencing the referee experience from initial recruitment through eventual departure from the role.

The themes that emerged at the Referee Recruitment stage were (a) *staying part of the game*, (b) *competition and challenge*, (c) *remuneration*, and (d) *socialization into the community*. All of these were positive benefits that worked concurrently to help attract and recruit the participants to their officiating roles (Warner et al., 2013). The authors suggested that these benefits should be highlighted to sport enthusiasts to help recruit more officials. As participants progressed to the next stage of Referee Retention, they began to encounter negative experiences that influenced their decisions to ultimately leave the role of officiating. Themes that emerged at the Referee Retention stage included (a) *problematic social interactions*, (b) *training/mentoring*, and (c) *lack of referee community*. According to the authors, the keys to retaining officials were to implement continued training and provide mentors to help referees deal more effectively with problematic social interactions with coaches, parents, and spectators. Additionally, Warner and colleagues emphasized the importance of fostering a sense of community and support among fellow referees.

At the final stage of Referee Advancement, three themes emerged, which included (a) *lack of administrator consideration*, (b) *administrator decision making*, and (c) *sport policies*. Participants who discontinued officiating did not feel supported or appreciated by league administrators and they perceived that referee assignments were based on favoritism and politics rather than on merit (Warner et al., 2013). Problems with career advancement were intensified when referees relocated or joined a different officials' association and encountered unstandardized policies. The authors suggested several strategies to increase retention at this stage that included more thoughtful consideration of individual needs by administrators, more transparent decision-making in regard to scheduling, and more standardized licensures for referees.

Overall, Warner et al. (2013) identified 10 themes or dimensions associated with referee development, and they underscored those that led to attrition. The findings of Warner et al. served as the foundation for the scale developed for this current study, as each of the themes they identified has been supported by both previous and subsequent studies of referees. A list of these dimensions, their descriptions, and supporting evidence for each of the factors can be found in Table 1. Warner et al. identified strategies to increase referee retention by preventing attrition. Because this study was comprehensive in terms of the number of factors (n=10) captured at recruitment, retention, and advancement stages, as well as its congruency with the extant refereeing literature, the dimensions outlined in Table 1 provided the starting point for this current study. The aim of the current study was to develop a tool to measure factors associated with the retention strategies suggested by the contemporary refereeing research (Cuskelly & Hoye, 2013; Kim, 2016; Kim & Hong, 2016; Phillips & Fairley, 2014; Ridinger, 2015; Schaeperkoetter, 2016; Tingle et al., 2014; Warner et al., 2013). Therefore, the purpose of this study was to develop a psychometrically sound scale to measure factors salient to referee retention.

<Insert Table 1 about here>

### **Methods & Results**

A survey instrument to measure referee retention was developed following the scale development guidelines suggested by Devellis (2012). These guidelines include the following eight steps: 1) determine clearly what it is you want to measure, 2) generate an item pool, 3) determine the format for measurement, 4) have initial item pool reviewed by experts, 5) consider inclusion of validation items, 6) administer items to a development sample, 7) evaluate the items, and 8) optimize scale length. Steps 1-5 were addressed during the first phase of this project,

which focused on instrument development. Next, during phase 2, a field test was conducted to address steps 6-8. Finally, in phase 3, the revised instrument was administered to a larger sample for the main study to confirm and validate the factor structure of the instrument. Details of each of these phases are described in the following sections.

### **Instrument Development (Phase 1)**

Determination of what is to be measured is the first step of instrument development (Devellis, 2012). This study sought to develop a tool to measure factors associated with referee retention. The 10 dimensions and corresponding retention strategies identified by Warner et al. (2013) served as the foundation to develop this tool. The second step of scale development involved generation of an item pool (Devellis, 2012). Based on the literature and officiating experience of two members of the research team, four to five items relevant to each of the 10 dimensions identified by Warner et al. (2013) were created. The items for each factor were designed to be similar since redundancy is useful and desired at this stage of scale development. As noted by Devellis (2012), “By using multiple and seemingly redundant items, the content that is common to the items will summate across items while their irrelevant idiosyncrasies will cancel out” (p. 78). Each item was reviewed and discussed by the research team until consensus on wording was reached. The third step of scale development was to determine the format for measurement (Devellis, 2012). A 5-point Likert-type scale ranging from 1 = strongly disagree to 5 = strongly agree was used to measure items comprising the 10 dimensions. A 5-point scale was utilized because it is the most commonly used scale in the social sciences and its familiarity with respondents could contribute to an increased response rate and response quality (Likert, 1932).

Next, Devellis' (2012) step four was implemented by having the initial item pool reviewed by experts. Eight individuals with advanced academic degrees and at least 10 years of officiating experience were selected as the panel of experts for this study. The group consisted of four women and four men with experience officiating different sports (i.e., field hockey, lacrosse, basketball, and football). Six of these individuals had master's degrees and two had doctorates and experience with scale development. Each panel member received a letter that explained the study's purpose and requested their assistance in establishing face and content validity by providing feedback about the survey and its specific items. A definition of each of the 10 dimensions was provided along with a copy of the survey instrument. The experts were asked to ascertain whether the items represented each construct by indicating the appropriateness of each item within its proposed dimension. In addition, the panel members were given several questions to guide their review. These questions asked about phrasing, terminology, clarity, and inquired if there were any specific items that should be added, deleted, or revised. Based on the experts' feedback, a number of revisions were made to the survey. For example, several items were reworded for greater clarity and consistency, and open-ended questions were added to the end of the survey. After the review by the panel of experts, a validation item was added per Devellis' (2012) scale development guidelines. For the field study, a 4-item measure of 'turnover intention' adapted from Abrams, Ando, and Hinkle (1998) was used for validation. For the main study, a 6-item measure of 'job satisfaction' adapted from Tsui, Egan, and O'Reilly (1992) was added for further validation.

### **Field Test (Phase 2)**

A field test was conducted as per steps 6-8 of Devellis' (2012) scale development guidelines. These steps included administering items to a development sample, evaluating the

items, and optimizing the scale length. An online survey instrument based on the results of phase 1 was created using Qualtrics software. Based on expert feedback, the survey included 38 items to measure the 10 dimensions identified by Warner et al. (2013) and four items to measure turnover intention (Abrams et al., 1998). Additionally, the survey included several demographic items and information about officiating experience, along with open-ended questions related to involvement with officiating.

### **Participants**

The participants for the field test included referees associated with the Amateur Athletic Union (AAU) Junior Olympics and individuals from local officiating associations known by the researchers. An email invitation with a link to the survey was sent by a representative of each group to a listserv of officials. The sample included 253 respondents. The AAU Junior Olympic officials were compared to the officials from the local associations on several items (i.e., age, gender, ethnicity, and years of officiating experience) and no significant differences were found so the groups were combined for further analysis. Most of the respondents were male (84%), Caucasian (78%) married (72%), and had a four-year college degree or higher (65%). The age of respondents ranged from 16-80 with a median age of 47. Years of officiating experience ranged from 1-53 with a median of 13 years. Respondents included referees from 13 different sports: primarily baseball (36%), track and field (20%), lacrosse (14%), wrestling (8%), and football (7%). Many of the respondents officiated at multiple levels including high school sports (84%), recreational youth teams (66%), middle school sports (65%), travel teams (56%), intercollegiate athletics (44%), college club teams (30%), Junior Olympic sports (30%), college intramurals (22%), and professional sports (10%).

## Data Analysis

First, an exploratory factor analysis (EFA) was conducted to explore the 10 factors based on the work of Warner et al. (2013) and to determine whether the factors were theoretically supported. EFA was performed with four analytical steps suggested by Thompson (2004) using the Statistical Program for the Social Sciences (SPSS) version 23.0.

- (1) Before conducting the EFA, a measure of sampling adequacy for EFA was examined using the Kaiser-Meyer-Olkin (KMO) statistic and Bartlett's test of sphericity. KMO examines whether items are sufficiently predicted for each factor while the Bartlett's test indicates whether items are highly correlated to provide a rationale for performing EFA. The KMO ranges from 0 to 1 and its acceptable value is greater than 0.7. Bartlett's test results should be significant ( $p < 0.05$ ).
- (2) Principal axis factoring (PAF) with varimax rotation was used to identify the factor structure. The number of factors was determined by using the Kaiser-Guttman rule (eigenvalues greater than 1) (Fabrigar, Wegener, MacCallum, & Strahan, 1999) and by inspecting the scree plot (the point of inflexion).
- (3) Factor loadings were inspected to elaborate the factor structure. Items were removed if they had factor loadings less than 0.4 or were conceptually incoherent within the factor. Multiple EFAs were conducted to finalize the factor structure while theoretical reasoning aided in the interpretation of each factor.
- (4) Reliability (internal consistency) of the final items was assessed by Cronbach's coefficient alpha. Cronbach's alpha values greater than 0.70 are considered as acceptable measures of internal consistency (Tabachnick & Fidell, 2001).

Second, a Rasch analysis was performed to investigate whether the five categories of the RRS properly functioned within the data. Rasch analysis has been determined to be a useful method to determine the optimal categorization of a rating scale (Myers et al., 2012; Zhu, 2001). A series of analytical steps were performed using WINSTEPS version 3.92.1 (Linacre, 2015). First, the rating scale was diagnosed based on Linacre's (2002) guidelines. This included the inspection of several items including the rating scale category probabilities, average measures order, outfit mean-square statistics, and step difficulty (threshold). Next, the category collapsing procedure was used to identify the most effective categorization structure.

### **Field Test Results**

#### **Measure of sampling adequacy.**

The results of the KMO statistic ( $KMO = 0.81$ ) and Bartlett's test of sphericity ( $\chi^2 = 3062.49$ ,  $df = 378$ ,  $p < 0.001$ ) indicated that the data were appropriate for factor analysis.

#### **The seven-factor Referee Retention Model.**

EFA results provided statistical support for a seven-factor model for the RRS. Based on the Kaiser-Guttman criterion, seven factors were identified that had eigenvalues greater than 1. In addition, the scree plot was examined to determine potential factors by identifying breaks in the slope within the plot. The results of the scree test supported the seven-factor solution. Inspection of factor loadings for the initial 38 items showed that 10 items had loadings less than 0.40 or were conceptually inconsistent with the emergent factors. Through the modification process, seven factors with 28 items were maintained for the final version of the RRS. Factor names and descriptions are displayed in Table 2 while factor loadings from the PAF with varimax rotation are shown in Table 3.

<Insert Table 2 about here>

The reduction of the original 10 dimensions into a seven-factor model resulted from items loading together and the elimination of one factor. Administrator Consideration was a composite of two items from Lack of Administrator Consideration and three items from Administrator Decision Making. Intrinsic Motives contained three items from Staying Part of the Game and three items from Competition and Challenge. Mentoring contained the same items as Socialization into the Officiating Community, but the new name of the factor (i.e., Mentoring) was deemed more appropriate since all items focused on the role of mentoring in the socialization process. All four of the original items for Remuneration were retained. Sense of Community included three of the five original items from Lack of Referee Community. The name of the factor was changed to Sense of Community, a term used more extensively in the literature to denote the feelings that group members have of belonging and being important to each other (Chavis, Hogge, McMillan, & Wandersman, 1986). Lack of Stress contained three of the four items from Problematic Social Interactions. The factor name was changed to Lack of Stress since the RRS is designed to measure factors associated with retention rather than attrition. Continuing Education was comprised of three of the five original items from Training/Mentoring. The three items that loaded on this factor dealt with ongoing training to prepare officials and thus, the factor was renamed to Continuing Education. Sport Policies, one of the 10 themes from Warner et al. (2013) was eliminated from the RRS because the items did not load on any of the emergent factors and these items only applied to a small sub-group of officials that had switched officiating associations. In an effort to make the scale more generalizable, the research team decided that the RRS should only retain factors that could apply to all officials.

<Insert Table 3 about here>

All of the items had acceptable factor loadings greater than 0.5 on their respective factor, thus fitting from a theoretical sense. Table 4 shows the eigenvalue associated with each factor, the percent of variance explained by each factor, the cumulative percent of variance explained, and the Cronbach's coefficient alpha for each factor. Administrator Consideration was the most important of the seven factors because it explained 22.28% of the total variation in officials' responses to the 28 scale items. Intrinsic Motives was identified as the second most important factor as it explained 11.74% of variation in officials' responses to the scale. The percentages of variance explained by other factors were: Mentoring (9.80%), Remuneration (8.32%), Sense of Community (6.17%), Lack of Stress (4.81%), and Continuing Education (4.18%). The seven factors accounted for 67.30% of the total variance in the items. The reliabilities (internal consistencies) for the seven factors ranged from 0.72 to 0.90, all within the acceptable range (Tabachnick & Fidell, 2001). The Cronbach's alpha of all 28 items combined was 0.83, which indicated good internal consistency.

<Insert Table 4 about here>

### **Category function of the initial RRS.**

Figure 1 shows the probability of responses for the five-category rating scale. The probability of response for Category 3 was below 0.50, indicating this category did not function well within the data.

<Insert Figure 1 about here>

A summary of the category functions of the RRS is reported in Table 5. The results demonstrate that the counts used per category were greater than 10 and there was a regular distribution (unimodal). Additionally, the average measure was considered acceptable as it advanced in order and the Outfit MnSq was less than 2.0. However, the category threshold (step

difficulty) was less than 1.0 logits between categories 3 and 4 suggesting the need for combined categories in the RRS. The five categories were collapsed into two different scales and compared. One scale included four categories and the other scale included three categories. The results suggested that the four-category rating scale was the most effective.

<Insert Table 5 about here>

### **Main Study (Phase 3)**

The refined RRS, which resulted from phase 2, was administered to a sample of 979 referees from two state-level high school athletic associations located in the mid-Atlantic region of the United States. Representatives from these two athletic associations sent an email invitation, with the survey link, to their respective members. The online survey included the refined 28-item RRS, demographic questions, and seven open-ended questions related to officiating. All survey responses were collected via the web-based Qualtrics system. Respondents from the two different athletic associations were compared on several items (i.e., age, gender, ethnicity, and years of officiating experience) and no significant differences were evident so the two groups were combined for further analysis.

### **Participants**

Of the 979 respondents, most were male (88%), Caucasian (79%), married (76%), and had a four-year college degree or higher (75%). The respondents' age ranged from 17-80 with a median age of 54. Years of officiating experience ranged from 1-51 with a median of 22 years. Respondents included referees from 13 different sports, with representation from basketball (26%), football (22%), soccer (14%), baseball (13%), softball (9%), volleyball (8%), wrestling (3%), lacrosse (3%), and field hockey (1%). Sports with less than 1% representation included track and field (0.5%), swimming and diving (0.4%), gymnastics (0.2%), and ice hockey (0.2%).

Many of the respondents officiated at multiple levels that included high school sports (94%), middle school sports (68%), recreational youth teams (63%), travel teams (50%), intercollegiate athletics (28%), college club teams (21%), college intramurals (13%), and professional sports (5%).

### **Data Analysis**

A confirmatory factor analysis (CFA) was conducted on the refined RRS items to evaluate the seven-factor model proposed by EFA results. The analysis was performed using Mplus software version 7.4. Model fit was evaluated using multiple indicators including the normed chi-square statistic ( $\chi^2/df$ ), the comparative fit index (CFI) (Hu & Bentler, 1999), the Tucker-Lewis index (TLI) (Tucker & Lewis, 1973), the root mean square error of approximation (RMSEA) (Schreiber, Nora, Stage, Barlow, & King, 2006), and the standardized root mean square residual (SRMR) (Bentler, 1990). The chi-square test evaluates the degree of a model's 'badness of fit' from the data where a non-significant p-value indicates a good model fit. However, because the  $\chi^2$  statistic is very dependent on sample size (Kline, 2015), the normed chi-square,  $\chi^2$  divided by the degrees of freedom ( $\chi^2/df$ ) was used to reduce the effect of sample size. The recommended values range from 0 to less than 5.0 (Schumacker & Lomax, 2010). Also, other goodness of fit indices were evaluated based on the following criteria: (a) the CFI and TLI values should be greater than 0.90, (b) the RMSEA should be less than 0.05, and (c) the SRMR values should be less than 0.05 to be an adequate model. Next, the four-category rating scale was evaluated for proper functioning using Rasch analysis (Linacre, 2002). The analytical steps followed the same procedures as reported in the field test.

To test convergent validity, associations between the RRS and the job satisfaction scale (Tsui et al., 1992) were analyzed using Pearson's correlation coefficients. This scale was selected

for a comparative measure because it is theoretically related to retention. The job satisfaction scale consisted of six questions about how satisfied one was with his or her work environment. Responses to a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree were used to calculate the mean scores. The indicators of strength of the Pearson's correlation coefficients are: 0.1 to 0.3 small; 0.3 to 0.5 medium; 0.5 to 1.0 large (Miles & Shevlin, 2001).

In order to provide evidence of predictive validity, multiple regression analysis was conducted with the turnover intention scale (Abrams et al., 1998) as the dependent variable and all seven factors of the RRS as independent variables. The turnover intention scale included four questions concerning plans to either leave or remain within an organization (Abrams et al., 1998). This analysis was done to determine if the RRS factors could be used as predictors of turnover intentions.

## **Main Study Results**

### **CFA results.**

The refined 28-item RRS scale with a seven factor-structure yielded an adequate measurement model for the data. Overall, CFA results indicated a good fit with the hypothesized model, with a normed chi-square of 2.87, a CFI of 0.930, a TLI of 0.920, and a RMSEA of 0.044 (90% CI: 0.040, 0.047). The SRMR was 0.055 which is at the boundary of being acceptable. These results provide empirical evidence that the seven-factor model sufficiently explained the item responses. Also, the standardized factor loadings of all items were statistically significant ( $p < 0.001$ ), ranging from 0.423 to 0.894. The standardized factor loadings and correlations among the factors are shown in Figure 2. The reliabilities (Cronbach's  $\alpha$ ) for the seven factors were: Administrator Consideration (0.89), Intrinsic Motives (0.80), Mentoring (0.87), Remuneration (0.70), Sense of Community (0.81), Lack of Stress (0.76), and Continuing Education (0.75).

<Insert Figure 2 about here>

### **Category function of the final Referee Retention Scale (RRS).**

The four-category rating scale function of the RRS is summarized in Table 6. Counts used for each category were greater than 10 and showed a regular distribution. In addition, the average measure advanced with each category, the Outfit MnSq was less than 2, and category thresholds increased by 1.0 logits. All of these results indicated the four-category RRS functioned well within the data.

<Insert Table 6 about here>

Figure 3 displays the probabilities of each category and shows its peak above 0.50 meaning that the usage of each category was appropriate. Overall, these results support the proper functioning of the refined four-category RRS.

<Insert Figure 3 about here>

### **Convergent validity**

The Pearson's correlation coefficient between the RRS and the job satisfaction scale demonstrated convergent validity of the scale. The RRS was positively associated with the job satisfaction scale with a strong correlation coefficient of 0.63 ( $p < 0.01$ ). This correlation can be seen as a sign of validity because job satisfaction and retention are theoretically comparable constructs.

### **Predictive validity**

The regression analyses revealed that factor 4 (Sense of Community) best predicted turnover intention,  $b = -0.31$ ,  $p < 0.001$ , followed by factor 2 (Intrinsic Motives),  $b = -0.24$ ,  $p < 0.001$ , and factor 5 (Lack of Stress),  $b = -0.15$ ,  $p < 0.001$ . The rest of the predictors were statistically insignificant. Combined, the seven factors explained a significant proportion of

variance in turnover intention,  $R^2 = 0.49$ ,  $F(7, 222) = 10.52$ ,  $p < 0.001$ . These results are displayed in the Table 7.

<Insert Table 7 about here>

### **Discussion**

Referees play a key role in the sport industry and there is continued concern about the growing shortage of qualified officials (American Sport Education Program, 2011; Cuskelly & Hoye, 2013; Kim, 2016). A better understanding of factors associated with referee retention can help sport administrators develop more effective recruitment and retention strategies for referees. While there are a few survey instruments specific to referees (Anshel et al., 2013; Kaissidis & Anshel, 1993; Myers et al., 2012; Taylor & Daniel, 1987), all of these scales focus on one specific psychological issue associated with officiating, either stress or self-efficacy. Recent literature suggests that in addition to psychological variables, there are also sociological (e.g., sense of community) and organizational (e.g., administrator consideration) factors that impact referee retention (Cuskelly & Hoye, 2013; Forbes & Livingston, 2013; Kim, 2016; Kellett & Warner, 2011; Ridinger, 2015; Tingle et al., 2014; Warner et al., 2013). As a result of this paradigm shift, much of the current research on sociological and organizational factors related to referee retention has been qualitative in nature, which is expected when little is known and a more in-depth understanding of a phenomenon is needed (Strauss & Corbin, 1998). As the sport officiating literature has evolved in the past decade, a more in-depth and multifaceted picture that includes psychological, sociological, and organizational factors of the referee experience at various career stages has been revealed.

While this more multidimensional knowledge of refereeing is useful, from a practical standpoint, the utility and its full impact has yet to be realized. Given that sport management is

an applied field, it is important to further this nascent line of research by providing sport leaders with the ability to predict referee retention. Such a tool will aid in referee retention by capturing the factors that need to be strengthened for a sport organization to better retain referees or for individual officials to pinpoint areas that need to be addressed for them to remain in the role. Further the RRS goes beyond previous scales that solely focused on the psychological issue associated with officiating, and also captures the organizational and sociological aspects while simultaneously considering various career stages (i.e., recruitment, retention, and advancement). To our knowledge, no extant tool considers a sport-focused approach and addresses the multifaceted nature of the referee experience. Thus, by utilizing previous sport-focused research that highlighted the multidimensional nature of refereeing, this study filled that gap by developing the RRS.

In an effort to contribute to literature and address a need to predict referee retention, an eight-step scale development procedure (Devellis, 2012) was followed to develop and test the RRS. This included a review of the scale by eight independent experts, an EFA conducted with field test data, and a CFA performed for the main study. A seven-factor, 28-item scale was determined to be a valid and reliable tool. Each individual factor, as well as the RRS as a whole, demonstrated strong internal consistency. The seven factors of the RRS combined to account for 67.30% of the total variance in responses to the 28 scale items. Administrator Consideration explained the most variance (22.28%) followed by Intrinsic Motives (11.74%). This finding suggests that a key retention factor is to develop procedures for assigning games that are fair and for administrators to consider the individual needs of each official. As numerous studies have identified administrative consideration as an important factor, this finding is not surprising (e.g., Kellett & Warner, 2011; Kim & Hong, 2016; Ridinger, 2015; Warner & Dixon, 2011). Also,

according to these findings, understanding the motives that connect referees to their role is important. Communication techniques to recruit and retain officials will resonate with individuals if the intrinsic motives revealed in this study are part of the message. These motives included: staying involved with the sport, giving back to the sport, and liking both the challenge of officiating and the competitive nature of sports. Furthermore, the importance of referees' motives has been clearly supported and highlighted by recent officiating research (e.g., Nordstrom et al., 2016; Phillips & Fairley, 2014; Schaeperkoetter, 2016).

Mentoring explained 9.80% of the variance and it is an important factor for initial attraction and integration into the officiating community. This finding supports numerous sport officiating researchers who specifically highlighted “mentoring” in their work (e.g., Kim & Hong, 2016; Nordstrom et al., 2016; Schaeperkoetter, 2016; Tingle et al., 2014). Less variance was explained by Remuneration (8.32%), Sense of Community (6.17%), Lack of Stress (4.81%), and Continuing Education (4.18%). Nevertheless, all of these factors contributed to the model and have been noted as important constructs in the literature (e.g., Kellett & Shilbury, 2007; Kellett & Warner, 2011; Rainey, 1995, 1999; Titlebaum, Haberlin, & Titlebaum, 2009; Warner et al., 2013).

A validity check of the RRS revealed a strong correlation (0.63,  $p < 0.01$ ) between the RRS and job satisfaction which was expected since research has shown that individuals with higher job satisfaction are less likely to depart from the profession than those with less job satisfaction (Eason, Mazerolle, Monsma, & Mensch, 2015). Since the RSS was correlated with job satisfaction and the literature provides strong evidence that job satisfaction is a core predictor of turnover intention (Boswell, Boudreau, & Tichy, 2005; Griffeth, Horn, & Gaertner, 2000; Tschopp, Grote, & Gerber, 2014), it was not surprising that several factors from the RSS were

found to be predictors of turnover intention. These predictors were Sense of Community ( $b = -0.31, p < 0.001$ ), Intrinsic Motives ( $b = -0.24, p < 0.001$ ), and Lack of Stress ( $b = -0.15, p < 0.001$ ). The inverse relationships indicate that high scores for each of these factors are associated with lower levels of turnover intention. Thus, retention strategies should focus on cultivating a strong sense of community, appealing to the intrinsic motives that attract individuals to officiating (i.e., enjoyment of competition and staying involved with a sport), and taking measures to reduce the occurrence of stressful situations for referees.

### **Conclusion**

With the growth of competitive sport, the significance of sport officials has become clear. The importance of understanding sport officials' experiences in terms of a more multidimensional view has been solidified in recent referee research. Specifically, the study by Warner et al. (2013) demonstrated the benefit of moving beyond the paradigm of studying referees exclusively through a human resource management or psychological lens. By applying the sport-specific theoretical framework of sport development (Green, 2005; Phillips & Fairley, 2014), a more inclusive picture of the referee experience was revealed (Warner et al, 2013). Indeed, important progress has been made through recent research (Forbes & Livingston, 2013; Kim & Hong, 2016; Nordstrom et al., 2016; Ridinger, 2015, Schaeperkoetter, 2016; Tingle et al., 2014; Warner et al., 2013), but these contributions and advancements fall short if they cannot be utilized by sport managers. This current work provides the practical application of the current body of knowledge regarding refereeing. The RRS, a 7 factor (28-item) survey, is a tool that can be used to evaluate and predict a referees' likelihood of retention. The results of administering the RRS will provide sport organizations with strategic information on how to better manage referees. More precisely, high scores on an identified factor (i.e., Administrator Consideration,

Intrinsic Motives, Mentoring, Remuneration, Sense of Community, Lack of Stress, or Continuing Education) will signify a strength associated with retention, while low scores on any one factor would indicate an area that needs improvement.

As with all research, there are limitations to note. The samples for the development of the RRS included AAU Junior Olympic officials and high school referees from a variety of different sports. Many of the respondents worked at multiple levels ranging from recreational youth leagues to professional sports. Nevertheless, the majority of respondents worked primarily at the high school and youth sport levels so the generalizability of the RRS to other levels of sport needs to be tested through future research. Furthermore, this study was limited to referees in the United States. Although sports officiating is a global phenomenon, there may be cultural nuances that make the officiating experience different in other parts of the world. Thus, it is recommended that referees from various countries are included in future studies utilizing the RRS.

Future research using the RRS will not only help narrow the gap between theory and practice (*cf.* Doherty, 2013), but it will also demonstrate the importance and distinctness of a sport-derived model (*cf.* Chalip, 2016). While this study and the findings are limited to the participants surveyed in two states, it does represent one of the most comprehensive and representative quantitative studies on referees since Rainey's work on referee abuse and stress in the 1990s (Rainey, 1995, 1999). Furthermore, the scale derived in this study highlights the important paradigm shift away from solely focusing on referee abuse and stress and is more inclusive of emergent literature and the current state of sport officiating knowledge (i.e., Cuskelly & Hoye, 2013; Kellett & Shilbury, 2007; Kellett & Warner, 2011; Kim, 2016; Nordstrom et al., 2016; Phillips & Fairley, 2014; Ridinger, 2015; Schaeperkoetter, 2016; Tingle

et. al, 2014; Warner et al., 2012, 2013). The RRS is a tool that can be used by researchers as well as practitioners to aid in the understanding of referee retention.

### References

- Abrams, D., Ando, K., and Hinkle, S. (1998) Psychological attachment to the group: Cross-cultural differences in organizational identification and subjective norms as predictors of workers' turnover intentions. *Personality and Social Psychology Bulletin*, *24*, 1027-1039. doi:10.1177/01461672982410001
- Anshel, M. H., Kang, M., & Jubenville, C. (2013). Sources of acute sport stress scale for sports officials: Rasch calibration. *Psychology of Sport and Exercise*, *14*, 362-370. doi: 10.1016/j.psychsport.2012.12.003
- American Sport Education Program. (2011). *Successful sports officiating* (2nd ed.). Champaign, IL: Human Kinetics.
- Barr, C. A., & Hums, M. A. (2012). Management principles applied to sport management. In L. P. Masteralexis, C. A. Barr, & M. A. Hums (Eds.), *Principles and Practice of Sport Management* (4th ed., pp. 65-44). Sudbury, MA: Jones & Bartlett Learning, LLC.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246. doi: 10.1037/0033-2909.107.2.238
- Boswell, W. R., Boudreau, J. W., & Tichy, J. (2005). The relationship between employee job change and job satisfaction: The honeymoon-hangover effect. *Journal of Applied Psychology*, *90*(5), 882-892.
- Chalip, L. (2006). Toward a distinctive sport management discipline. *Journal of Sport Management*, *20*, 1-21. doi:10.1123/jsm.20.1.1
- Chavis, D. M., Hogge, J. H., McMillan, D. W., & Wandersman, A. (1986). Sense of community through Brunswick's lens: A first look. *American Journal of Community Psychology*, *14*, 24-40.

- Chelladurai, P. (2014). *Managing organizations for sport and physical activity* (4th ed.). Scottsdale, AZ: Holcomb Hathaway Publishers.
- Cuskelly, G., & Hoye, R. (2004). *Problems and issues in recruitment and retention of sports officials. A report prepared for the Australian Sports Commission*. Brisbane Queensland: Griffith University.
- Cuskelly, G., & Hoye, R. (2013). Sports officials' intention to continue. *Sport Management Review, 16*, 451-464. doi: 10.1016/j.smr.2013.01.003
- Devellis, R. F., (2012). *Scale development theory and practice* (3rd ed.). Los Angeles, CA: Sage Publications, Inc.
- Doherty, A. (2013). Investing in sport management: The value of good theory. *Sport Management Review, 16*, 5-11. doi: 10.1016/j.smr.2011.12.006
- Eason, C. M., Mazerolle, S. M., Monsma, E. V., & Mensch, J. M. (2015). The role of personality in job satisfaction among collegiate athletic trainers. *Journal of Athletic Training, 50*(12), 1247-1255. doi: 10.4085/1062-6050-50.11.08
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272–299. doi: 10.1037/1082-989X.4.3.272
- Forbes, S. L., & Livingston, L. A. (2013). Changing the call: Rethinking attrition and retention in the ice hockey officiating ranks. *Sport in Society, 16*(3), 295-309. doi: 10.1080/17430437.2013.779854
- Furst, D. M. (1989). Sport role socialization: Initial entry into the subculture of officiating. *Journal of Sport Behavior, 12*(1), 41-52.
- Furst, D. M., (1991). Career contingencies: Patterns of initial entry and continuity in collegiate

- sports officiating. *Journal of Sport Behavior*, 12(2), 93-102.
- Green, C. (2005). Building sport programs to optimize athlete recruitment, retention, and transition: Toward a normative theory of sport development. *Journal of Sport Management*, 19, 233-253. doi: 10.1123/jsm.19.3.233
- Griffeth, R. W., Horn, P. W., & Gaertner, S. A. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26, 463-488.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. doi: 10.1080/10705519909540118
- Infographic: The nationwide decline in sports officials. (2016). *Athletic Business*. Retrieved from <http://www.athleticbusiness.com/recreation/infographic-the-nationwide-decline-in-sports-officials.html>
- Kaissidis, A., & Anshel, M. H. (1993). Sources and intensity of acute stress in adolescent and adult Australian basketball referees: A preliminary study. *Australian Journal of Science and Medicine in Sport*, 25, 97-103.
- Kellett, P., & Shilbury, D. (2007). Umpire participation: Is abuse really the issue? *Sport Management Review*, 10, 209-229. doi: 10.1016/S1441-3523(07)70012-8
- Kellett, P., & Warner, S. (2011). Creating communities that lead to retention: The social worlds and communities of umpires. *European Sport Management Quarterly*, 11(5), 471-494. doi: 10.1080/16184742.2011.624109
- Kim, S. (2016). Perceived organizational support as a mediator between distributive justice and sports referees' job satisfaction and career commitment. *Annals of Leisure Research*,

doi: 10.1080/11745398.2016.1147363

- Kim, M. C., & Hong, E. (2016) A red card for women: Female officials ostracized in South Korean football. *Asian Journal of Women's Studies*, 22, 114-130. doi: 10.1080/12259276.2016.1168156
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY: Guilford publications.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology* 140, 1-55.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2015). WINSTEPS Rasch measurement computer program (Version 3.92.1). Chicago, IL: Winsteps.com.
- Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. Thousand Oaks, CA: Sage.
- Myers, N. D., Feltz, D. L., Guillen, F., & Dithurbide, L. (2012). Development of, and initial validity evidence for, the Referee Self-Efficacy Scale: A multistudy report. *Journal of Sport & Exercise*, 34, 737-765. doi: [10.1123/jsep.34.6.737](https://doi.org/10.1123/jsep.34.6.737)
- Nordstrom, H., Warner, S., & Barnes, J. (2016). Behind the stripes: Female football officials' experiences. *International Journal of Sport Management and Marketing*, 16, 259-279. doi: 10.1504/IJSMM.2016.077934
- Phillips, P., & Fairley, S. (2014). Umpiring: A serious leisure choice. *Journal of Leisure Research*, 46(2), 184-202.

- Purdy, D. A., & Snyder, E. E. (1985). A social profile of high school basketball officials. *Journal of Sport Behavior*, 8(1), 54-65.
- Rainey, D. (1995). Sources of stress among baseball and softball umpires. *Journal of Applied Sport Psychology*, 7, 1-10. doi: 10.1080/10413209508406296
- Rainey, D. (1999). Sources of stress, burnout, and intention to terminate among basketball referees. *Journal of Sport Behavior*, 22, 578-588.
- Rainey, D. W., & Hardy, L. (1999). Sources of stress, burnout and intention to terminate among Rugby Union referees. *Journal of Sports Sciences*, 17, 797-806. doi: 10.1080/026404199365515
- Read, D. (2000). Stars in stripes. *Athletic Management*, 22(6), 26-32.
- Ridinger, L. L. (2015). Contributors and constraints to involvement with youth sport officiating. *Journal of Amateur Sport*, 1(2), 103-127. doi: [10.17161/jas.v1i2.4946](https://doi.org/10.17161/jas.v1i2.4946)
- Schaeperkoetter, C. C. (2016). Basketball officiating as a gendered arena: An autoethnography. *Sport Management Review*. doi: 10.1016/j.smr.2016.05.001
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338. doi: 10.3200/JOER.99.6.323-338
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York: Routledge.
- Stevens, T. (2016). Abusive fans make it tougher to recruit high school sports refs. *The News Observer*. Retrieved from [www.newsobserver.com/sports/high-school/article68495447.html](http://www.newsobserver.com/sports/high-school/article68495447.html)

- Strauss, A., & Corbin, J. C. (1998). *Basics of qualitative research techniques and procedures for developing grounded theory* (2nd ed.). London, Sage Publications.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.
- Taylor, A. H., & Daniel, J. V. (1987). Sources of stress in soccer officiating: An empirical study. In T. Reilly, A. Lees, K. Davids, & W.J. Murphy *Science and Football: Proceedings of the First World Congress of Science and Football*. London: E & FN Spon.
- Taylor, A. H., Daniel, J. V., Leith, L., & Burke, R. J. (1990). Perceived stress, psychological burnout and paths to turnover intentions among sport officials. *Applied Sport Psychology*, 2, 84-97. doi: 10.1080/10413209008406422
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association. doi: 10.1037/10694-000
- Tingle, J. K., Warner, S., & Sartore-Baldwin, M. L. (2014). The experience of former women officials and the impact on the sporting community. *Sex Roles*, 71, 7-20. doi: 10.1007/s11199-014-0366-8
- Titlebaum, P. J., Haberlin, N., & Titlebaum, G. (2009). Recruitment and retention of sports officials. *Recreational Sports Journal*, 33(2). doi: 10.1123/rsj.33.2.102
- Topp, B. (2001, May). Sports officials becoming endangered species. *The National Association of Sports Officials*. Retrieved from <http://www.naso.org/Research/st80.htm>
- Tschopp, C., Grote, G., & Gerber, M. (2014). How career orientation shapes the job satisfaction-turnover intention link. *Journal of Organizational Behavior*, 35, 151-171. doi: 10.1002/job.1857

- Tsui, A. S., Egan, T. D., & O'Reilly, C. A. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly*, 37(4) 549-579. doi: 10.2307/2393472
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10. doi: 10.1007/BF02291170
- Warner, S., & Dixon, M. A. (2011). Understanding sense of community from an athlete's perspective. *Journal of Sport Management*, 25, 258-272. doi: [10.1123/jsm.25.3.257](https://doi.org/10.1123/jsm.25.3.257)
- Warner, S., Tingle, J. K., & Kellett, P. (2013). Officiating attrition: The experiences of former referees via a sport development lens. *Journal of Sport Management*, 27(4), 316-328. doi: 10.1123/jsm.27.4.316
- Zhu, W. (2001). A confirmatory study of Rasch-based optimal categorization of a rating scale. *Journal of Applied measurement*, 3(1), 1-15.