The Development of an Individuals-within-Dyads Multilevel Performance Measure for an Interactive Cheerleading Task

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Measurement in Physical Education and Exercise Science on 19 Oct 2015, available online: http://www.tandfonline.com/10.1080/1091367X.2015.1082474

Abstract

Dyadic interactions generate direct relationships in which interdependent sport behaviors can be destructured. The focus of this investigation was to develop a two-level performance framework and corresponding measures of individual- and dyad-level sport performance. The described procedures surrounded a male-female cheerleading paired-stunt task, as only team-level outcomes are currently assessed during sport competition. Multiple observers employed the developed measures (α = .89 - .96; ICC = .87 - .95) to assess the videoed performance of 132 individuals nested within 66 intact dyads competing at a national competition. Unique information is revealed from each partner's individual-level score, disjointedly assessed, and their dyad-level score, an assessment of combined efforts. Score differences are especially apparent when in contrast to an aggregated dyad-level score. A discussion of the outlined approach and interpretation of multilevel occurrences of interdependent processes and outcomes of sporting performance is provided.

Keywords: Dyad, Individual, Performance, Measurement, Interdependence

The Development of an Individuals-within-Dyads Multilevel Performance Measure for an Interactive Cheerleading Task

Adequate team performance, typically, cannot be achieved without each individual performing his or her role to some degree of correctness. However, team performance is not always equivalent to the sum of individual parts and the quality of *interactions* between team members is often influential on success. Team statistics (e.g., goals, turnovers, assists, pass completions, etc.) are typically used as indicators of performance, but the critical analysis of individual contributions to the interactive components of performance could augment knowledge about the success of winning teams (Fernandez, Camerino, Anguera, & Jonsson, 2009). Unfortunately, little research has been directed towards analyzing interdependent skills in team sport performance (Hughes & Bartlett, 2002; Travassos, Davids, Araújo, & Esteves, 2013).

Subgroups of differing size will exist within a team. However, a *dyad* ties the individual to the group in the simplest form. The dyad is the only size of group that cannot be further divided into subgroups (Levine & Moreland, 1998). In addition, two-person interactions are substantially void of third-party behavioral influences (Levine & Mooreland, 2006). Larger group sizes generate a more complex network of dynamic and mutual influences that must be accounted for among individuals (Hare, 1976). Therefore, dyads function as an elementary unit in which to understand and measure individual performance within any size sporting group (McGarry, 2009). Despite this, interpersonal behavior studies within sport tend to focus on larger size teams (Gaudreau, Fecteau, & Perreault, 2010).

Organizing team-level behaviors into lower-level dyadic interactions also elicits consideration of the individuals within the dyads. Researchers conceptually approach relationships as two partners acting as interdependent units with distinct contributions towards developmental outcomes (Laursen & Bukowski, 1997). That is, two individuals will have multiple, bidirectional interconnections through simultaneously producing and responding to one another's behaviors (Laursen, 2005; Malloy & Albright, 2001). However, without performance information related to each individual disjointedly, the mutual influences can only be assumed (Kenny, Kashy, & Cook, 2006). A unidirectional measure of one partner's performance does not provide adequate information about the other partner's performance, or their performance together, for that matter. Laursen (2005) further clarifies that data from each partner is more revealing of interdependent behaviors because, in general, measuring variance within relationships is impractical if only one source of variation is addressed and reported.

Conceptually, team accomplishments derive from the individuals that make up the team *and* the interactions between those individuals (Arthur et al., 2005; Tesluk, Mathieu, & Zaccaro, 1997). Multiple points of data from varying levels and perspectives, as an adequate representation of two interacting performers, include (a) a measure of partner A's individual performance, (b) a measure of partner B's individual performance, and arguably (c) a measure of their performance together at the dyad level. While data points (a) and (b) are theoretically nested under data point (c), an aggregation score may be statistically misleading and conceptually meaningless for a dyad-level performance score (Malloy & Albright, 2001). As illustration, the interaction of two moderate performers and the interaction of one great performer and one poor performer are not analogous interactions but could be represented by identical mean scores (Laursen, 2005). Not only could the sum of points (a) and (b) be unknowingly incongruent to data point (c), but the uniqueness within each dyadic interaction, as determined by the individuals within those interactions, is removed from a dyad- level performance score. The complexity of a team performance decomposed into dyadic interactions that are further decomposed into individual behaviors provides one approach to understanding how a particular performance, at any given level, occurs in the context of team competitions (Travassos et al., 2013).

The aim of this investigation was to identify, measure, and describe aspects of individual and interactive outcomes within joint sport performance conditions. First, we provided a conceptual breakdown of a performance task into two contributing levels and individual role requirements. Second, an adequate measure of each theoretically contributing unit (i.e., partner A, partner B, and dyad) within task performance was developed. Relatedly, initial validity and reliability of each measure was determined using a panel of observers. Third, relationships among performance scores were analyzed so as to interpret the reality of individuals performing within a conjoint performance outcome. The current investigation provides implications for improving sport scoring procedures, a systematic framework of task outcome quality from multiple perspectives, and descriptive relationships of real dyadic performance data.

Sport Specific Background

Like many aesthetic sports, cheerleading performances are given a score linked to task difficulty and points are deducted for differences between the desired behavior and observed behavior (i.e., errors; Zelaznik, 2014). A brief review of scoring systems in interdependent paired sports similar to cheerleading (i.e., synchronized diving, synchronized swimming, paired acrobatic gymnastics, paired ice-skating, and paired dance) revealed wide variability in assessment quantification conventions. Each sport scoring procedure was unique and, therefore, did not provide a shared fundamental approach to measuring dyadic sport performance. However, performance was commonly indicated by both individual- and dyad-level aspects across each scoring system. In comparison, competitive cheerleading scores are awarded for the overall quality of team execution even though many of the tasks are completed by subgroups. Therefore, the focus of this investigation was to develop measures for the performance of individuals and dyads within cheerleading teams.

Research questions that address lower-level contributions toward a team outcome require concordant outcome measures (Kenny, Mannetti, Pierro, Livi, & Kashy, 2002). For this investigation, the judges' team-level scores are less valuable because variability attributed to dyads and individuals is absent (Laursen, 2005; Tesluk et al., 1997). Judges presumably assess the individuals and dyads in some sense in evaluating the team, but there is not a written score that represents or provides direct evidence for individual or dyad performance assessments. The current investigation is not intended to be used for criticism of the scoring system within competitive cheerleading. The investigation may provide a unique account of lower-level contributions often ignored in highly interdependent team sports.

Measures in this investigation were developed in relation to one interdependent paired-stunt task performed within each team routine. Successful task completion required multiple sets of male-female dyads to perform the same paired-task in unison. The task was chosen because it provided observable differences between individual-, dyad-, and team-level performance requirements. Within the context of this investigation, individual-level performance was defined as utilizing proper technique required by one's unique role. The dyad-level performance was defined as the degree to which two partners integrate behaviors so as to avoid errors, and the team-level performance was defined as the quality of unified movements across dyads (although not assessed in this investigation).

Performance Modelling and Observation Framework

Performance assessment interests researchers, practitioners, coaches, judges, and athletes (O'Donoghue, 2010). To suit the range of persons who would find these measures useful, a scientific approach and coaching perspective were merged (Franks & Goodman, 1986). This required considerable knowledge in the sport as well as a review of existent performance models and observation schemes used in more traditional team sports.

A theoretical definition of performance is a fundamental starting point for the development of a performance measurement scheme (Morrison, 2000). With the idea that performance is an execution of action, at least in aesthetic sports, many accept the terms technique and performance as synonymous. However, the two terms are not equivalent (Lees, 2002). Better performance outcomes do not directly indicate a better use of technique. For example, observing the landing of a tumbling skill does not indicate the form used in-flight directly before the landing. Thus, one cannot assume that the completion of the task (performance outcome) is equivalent to the aesthetic quality of the acrobatic skill (proper execution or technique; Hauw, Renault, & Durand, 2008). However, observed first-rate technique does tend to indicate a better performance quality and outcome. Therefore, analysis of technique, a process-oriented rather than outcome-oriented analysis, is a superior indicator of performance compared to performance outcomes alone (Barnett et al., 2009; Burton & Miller, 1998).

Technique is a sequence of movements described as body lines and angles in relation to their temporal occurrences (Lees, 2002). Technique analysis is a common concept that lacks a strong conceptual framework partly due to the unique requirements of every sport skill, especially those of acrobatic nature. While the biomechanics of a movement pattern may be valuable for technique analysis, the detailed process often lacks the capability to meaningfully link pieces of information to an entire task (Lees, 2002). Qualitative technique analyses, as characterized by subjective observations of performance, are more common in applied settings. While qualitative observation usually requires extensive knowledge about the task, it is relatively less time consuming and can be used by people with a varying range of expertise (Lees, 2002). Qualitative procedures best addressed the purposes of this investigation because, within performance settings, analyses regarding whole-body movement are likely more warranted and applicable by coaches and judges. Intricate details are a partial representation of performance and often require supplementary qualitative considerations to make sense of the details (Hughes & Bartlett, 2002).

McPherson's (1990) and Hay and Reid's (1982, 1988) models for qualitative movement diagnosis both emphasize that creating a systematic observation scheme to accurately detect errors relies on heavy inquiry during two steps within the preobservation stage. Step one, *movement analysis*, includes the identification of critical features of the skill and consideration of the factors that alter perception of the skill. Step two, *planning the observation*, includes developing a recording form and outlining an assessment process (McPherson, 1990). The manner in which important features of the skill are highlighted will prompt the observation scheme, organization of assessment, and recording tool.

Movement Analysis. The organization of an observation scheme is important because it will directly influence how a movement is perceived (Knudson, 2013). Gangstead and Beveridge's (1984) model for sport skill observation and analysis is a well-used qualitative model. The systematic observation model operates to indicate discrepancies in actual and desired behaviors across multiple athletes while reducing the complex visual display of a body in motion (Gangstead & Beveridge, 1984). Drawing observer attention to specific parts of movement through spatial and temporal markers so as to reduce the observer's perceptual load is a unique feature of the model.

Gangstead and Beveridge (1984) stated their model was fashioned to manipulate the observers' visual experiences so as to simplify the evaluation process. Lees' (2002) review of technique analysis in sport highlights three strategies to organize observational depictions of movement as in line with Gangstead and Beveridge (1984). First, *phase analysis* involves breaking down a task into subjective phases of movement determined by the specific task and purpose of analyses. Second, *temporal analysis* is related to the rhythm, timing, and the sequences of movements important to performance. Noticeably, phase and temporal analyses are intuitively intertwined. Third, *critical features*, or components absolutely essential to the skill, are identified. Critical features must be observable aspects, least modified by a performer, to achieve the desired outcome. The most common strategy used to control observer perceptual load has been to indicate critical features within temporal phases (Gangstead & Beveridge, 1984; James & Dufek, 1993; Lees, 2002).

Planning the Observation. To aid observation and evaluation, model performance templates are created to provide observers with ideal representations of movements. Any deviations from the model template characterize quality of technique within performance. Templates are multilayered and generalizable across many athletes, yet specifically vary according to how the skill is subjectively analyzed (Knudson, 2013). When skills are more complex, increasing the number of observation trials is argued to relieve limited perceptual capacities of the observers (Knudson, 2013). Hay and Reid (1988) suggest two or three observation trials should be utilized for gathering a general impression of the movements, and then further trials can be focused on parts of movement for systematic review. Additional strategies suggested to increase the validity and reliability of subjective observations include providing observer training, specifying measurement guidelines, and simply increasing the total number of trials (Knudson, 2013).

Investigative Rationale

Team sports including net and wall games, invasion games, and striking and fielding games, have been provided with recommended performance indicators for analytical purposes (Hughes & Bartlett, 2002). The current investigation, to our knowledge, was the first attempt to provide performance indicators for dyadic relationships within aesthetically-based, interdependent team sports. This involved indicating a logical structure of interdependent performance relative to the specific sport yet grounded in previous performance analysis models. Measurement scales were then developed and applied in a performance assessment. As a consequence, each paired-stunt task produced three performance scores representing (a) the male's performance, (b) the female's performance, and (c) their conjoint performance.

Relationships between differing outcome scores within the performance framework were expected. The dyad-level performance score was hypothesized to relate positively with both individual-level scores due to the conceptual deployment of the performance hierarchy. In addition, the individual-level performance scores were hypothesized relate positively with one another because the scores were measured at the same level. To further clarify how each level of analysis related, an aggregated dyad-level score was created. Aggregating individuals' scores to indicate group-level variables is a common practice associated with many statistical and conceptual issues (Tesluk et al., 1997). Relationships were compared between the three observed performance scores and the aggregated dyad-level score to illustrate the extent to which a purposefully designed dyad measure uniquely informed conjoint performance. In addition, evidence of observer agreement was to be identified for each of the performance scales. Finally, interpretations of performance scores were aimed at emphasizing the natural interdependencies within a competing sport team. The unique attributes of the paired-stunt task caused a largely exploratory nature of task analysis. However, the clear divisions between levels of measurement and the divided contributions from each individual provided an ideal structure for this investigation to operate within.

Method

Participants

Sixty-six cheerleading dyads (132 individuals) competing in a national competition agreed to participate in a larger study. Each athlete performed within only one dyad which comprised of one female (flyer role) and one male (base role).[1] Eleven university teams from the southwest (*n* = 5), southeast (*n* = 3), midwest (*n* = 2), and west (*n* = 1) regions within the United States were included in the study. Participants were from 18 to 31 years of age (base

M = 22.13 years, *SD* = 2.93; flyer *M* = 19.88 years, *SD* = 1.45). The larger study was approved by the institutional review board and the informed consent explicitly acknowledged that participation included assessment of publicly-available video recordings of competition routines described in this investigation. No additional video recording occurred.

Performance Task

The paired-stunt task had an average duration of 6.5 s (*SD* = 1.86, range = 6 – 11 s) within a two and a half minute competition routine.[2] Task inclusion criteria specified that performance of task skills would require only one base and one flyer. Safety rules implemented by the competition required very difficult skills to be performed with an additional spotter. To eliminate any third-person confounds, task exclusion criteria dismissed any skills requiring a spotter. As a consequence of the inclusion and exclusion criteria, tasks were comparable across teams. All tasks followed a sequence of the flyer being freely tossed from her hips into the air by the base (entrance). The flyer's feet then land on the base's hands that are formed as a platform. The flyer holds a controlled body position while being extended in the air (middle portion). For the termination of the skill, the base releases the flyer's feet and catches the flyer's hips to assist her two-footed landing on the ground.

Performance Measures

Dyad-level Performance. The assessment of dyad-level performance was adapted from the National Cheerleading Association's (NCA; 2013) college rulebook. The gold-standard scoring system provided guidelines, originally created for team performance assessments, applicable to each dyad. Both *difficulty* and *execution* were components of the dyad scores. Dyad-level performance ranged from 0 - 10 with higher scores representing proper execution (less errors) of a more elite skill range. Categories of skills were provided from the NCA (2013) rulebook and are listed in Table 1. Each dyad was placed in a score range linked to the difficulty of the attempted performance task and deductions of 0 - 2 points, in increments of

.5 points, were given in accordance to gradations of errors (NCA, 2013). Table 2 provides descriptions for the appropriate allocation of deductions. For analyses, each dyad-level performance score was the mean of four observers' scores.

Individual-level Performance. Each individual was assessed on nine dimensions. The athlete's body was divided into three segments; *arms and shoulders*, *core and hips*, and *legs and feet*. Each body segment was then assessed across three temporal phases of the performance task; *entrance*, *middle portion*, and *dismount*. The nine dimensions were each assessed on a four-point, Likert-type scale and then summed. The four-point scale included anchors at 0 (*perfect technique/no errors*), -1 (*imperfections/flaws*), -2 (*mistakes/slip ups*), and -3 (*failures/unsuccessful*). The lowest total score (-27) indicated poor performance (numerous errors) and the highest total score (0) indicated excellent performance (no errors). For analyses, each individual-level performance was the mean of two observers' total scores.

Critical features were identified for individual performance relative to the unique performance requirements of each role. The execution of technique was expressed in four levels of quality across each dimension in concordance to the Likert-scale anchors resulting in 27 portrayals of possible movement features for each role. For example, the following critical features refer to the legs and feet during the middle portion of the task. For the base, points were deducted as follows: (0) legs absorb as needed, placement of legs should be just outside of shoulders with knees forward, staying in same spot, (-1) stance is too wide or narrow, legs are not utilized to absorb, one step, (-2) small, unnecessary steps are taken and stunt remains in air, (-3) lots of unnecessary steps are taken, does not save stunt. Separate critical features indicating decrements of role-specific errors were identified for the flyers. [3]

Procedure

The procedures were mostly directed towards development of the individual-level scoring system as the dyad-level scoring system was an adaption of an existent performance scoring procedure employed within the sport. However, both systems were piloted through observer training and adjustments were made accordingly.

Movement Analysis. For movement analysis, critical features and factors altering perception required consideration. Input from a panel of university-level coaches provided information regarding how performance is typically perceived by experts. Six current college coaches (five males and one female) were each asked to list the five most important aspects for successful performance by the base and flyer separately. Answers from the coaches provided several links to observation strategies in the existent literature. Specifically, aspects were uniquely identified for only parts of the task reflecting the temporal phases approach.

Coaches also made apparent the complexity of the acrobatic movements. Not only were the critical features different for the base and flyer, but they were associated to specific sections of the body. Therefore, observer attention would need to be specifically directed towards general moving parts so as to reduce perceptual load. Dividing the body into three segments structured across each temporal phase of the task reduced the degrees of freedom and kept the partners' performances intuitively connected. Critical features for each dimension were developed by the first author (first author has cheerleading experience as a top-level performer and coach of top-level athletes) and edited several times after conferring with the same coaches and later with observers when issues arose.

Planning the Observation. Planning the observation required attention towards outlining the process of assessment and developing a suitable recording form. Four current co-ed college cheerleaders (two males and two females) were recruited as adept performance observers in this investigation. Each observer had an average of four years of experience in co-ed style cheerleading, an average of nine years of overall experience in cheerleading, at least three years of experience competing at the national level, and each had participated in various judging opportunities within the sport.

The observers received three training sessions, each lasting about two hours. In the first training session, the observers were introduced to the critical features for both roles that would guide performance assessment. Following an explanation of the critical features, the four observers and first author discussed potentially confusing issues as well as possibilities not considered. The following training sessions required the observers to practice assessing individuals and dyads from random teams not participating in the investigation.

During the second observer training session, the four observers assessed all dyad and all individual performances for a random team (six dyad performances with the associated 12 individual performances per observer). Perhaps unsurprisingly in retrospect, it became obvious that role-experience had an influence on the observers' ability to score individual performance in those roles. Perceptual abilities linked to observer experience have long been known to be an overriding issue of technique analysis (Armstrong & Hoffman, 1979; Biscan & Hoffman, 1976; Weekley & Gier, 1989), so responsibility for observing individual performances was divided with the male observers assessing the base performances and the female observers assessing the flyer performances. This division of individual-level observation was piloted during the third observer training session (10 dyad performances with associated 10 individual performances per observer). Both observers and first author were satisfied with the reduced workload and increased observer confidence to employ the rating scales, so this served as the final training session for the observers. Viewing sessions during data collection were operated in correspondence to the final training session.

Performance Assessment Protocol. Professional videos by competition personnel were made available on the internet and projected onto a large screen. No more than 17 dyads (about two teams) were observed in one sitting with a maximum of two hours per observation session. Each dyad score required about 60 - 90 s to complete (three trials) and each individual score required about three to four minutes (10 trials) to complete. Observations were completed in five sessions, resulting in about 10 hours of performance assessment per observer (3,372 trials in sum).

Across Sessions. To reduce measurement errors attributable to observers' varying attitudes, effort, and emotions across the five viewing sessions, a warm-up dyad (randomly chosen from a nonparticipating team) was observed. All dyads included in the investigation were randomly assigned to a viewing order within their respective team and all teams were viewed in a random order. Within each team, individual performances were assessed after, and in the same random order as, the dyads. For all viewings, the entire team was visible to the observers on the projected screen. The video was played a few seconds before the task and stopped immediately after the task was completed. Observers were never able to see other skill-elements performed in the routines. Before each trial, the dyad of focus was indicated with a red laser-pointer and verbalized by the first author to direct attention towards where the dyad of interest would begin the task.

Within Sessions. In line with Hay and Reid's (1988) recommendations, observation trials were provided for both general impressions and more specific parts of movements. For each team, observers first watched performance without focus on a particular dyad or individual and determined the performed task's difficulty range from the provided dyad-level scoring guidelines. After which, the observers had three trials to assess each dyad-level performance. First, the observers familiarized to the pair's general movements and were asked to not record any values. During the next two trials, the observers assessed a starting score within the appropriate range, designated any deductions, and recorded the final score.

Individual-level performances were assessed by watching ten trials of each participant. As before, the observers were given one trial to familiarize with the general movements of the particular individual he or she was observing without recording any values. For the remaining trials, observers were asked to utilize three trials per a body segment; always beginning with the arms and shoulders and ending with the legs and feet. The body segment order was fixed, but the observers were given freedom to appraise technique during the temporal phases as willed. The limited freedom reduced perceptual load, maximized knowledge of the entire task in relation to a specific body section, and reinforced utilizing all trials to provide an accurate performance score. Before each trial, the body segment of focus and number of remaining trials were verbalized. Between trials, all observers utilized the critical features to assign performance scores as all written guidelines were readily accessible.

Analysis

Analyses were conducted using SPSS (version 21.0) to examine the reliability of the two related performance measures. Cronbach's alphas and interclass correlations were calculated to observe the relatedness among observations as well as the consistency of each observer. Pearson product-moment correlations were calculated as an indication of how the performances were generally related. Furthermore, a dyad-level aggregation score was generated, from the base and flyer scores, to demonstrate the potential differences in the type of score representing a dyad's performance.

Results

The flyers' performance scores ranged from -20.5 to -3 points (*M* = -8.39, *SD* = 3.74). The bases' performance scores ranged from -23.5 to -4.5 points (*M* = -12.80, *SD* = 3.94). Performance scores for flyers were non-normally distributed with skewness of -1.27 (*SE* = 0.30) and kurtosis of 1.98 (*SE* = 0.58). Performance scores for bases were more normally distributed with skewness of -0.55 (*SE* = 0.30) and kurtosis of 0.40 (*SE* = 0.58). Dyad-level performance ranged from 3.72 to 8.78 points (*M* = 6.97, *SD* = 1.06). Normality was more similar to the base performance distribution as the skewness value was -0.58 (*SE* = 0.30) and the kurtosis value -0.06 (*SE* = 0.58) for dyadic performance scores. The aggregated dyad performance scores ranged from -22 to -5.25 (*M* = -10.50, *SD* = 3.50) with skewness of -1.11

(*SE* = 0.30) and kurtosis of 1.24 (*SE* = 0.58).

Within this sample, the performance scores for flyer (α = .89) and base (α = .89) were provided from two observers while the performance scores for the dyad (α = .96) were provided from four observers. Interclass correlations set to absolute agreement for the base (.87, $p < .001$, 95% *CI* [0.76, 0.93]), flyer (.88, $p < .001$, 95% *CI* [0.79, 0.93]), and dyad (.95, $p < .001$, 95% *CI* [0.92, 0.97]) performance scores with average measures were observed for the same observer groups.

The flyer and base performance scores, assessed by independent observer pairs, had a moderately high correlation ($r = .69$, $p < .01$). This indicates if either the flyer or base performed well (committed less errors) then his or her partner would likely have also performed well. The relationships between the dyad-level score and each individual-level role score were both positive and moderate. The flyer performance scores had only a slightly stronger relationship ($r = .42$, $p < .01$) than the base performance scores ($r = .35$, $p < .01$) with the dyad scores. This indicates neither role was dominantly related to the dyad performance scores. The aggregated dyad scores were only moderately related to the observed dyad-level scores ($r = .42$, $p < .01$) indicating that the two dyad-level indices are not equivalent. As expected in an aggregation index, both the base ($r = .92$, $p < .001$) and flyer ($r = .91$, $p < .001$) individual scores were almost perfectly associated with the aggregated dyad scores. This further highlights that relationships between individual- and dyad-level scores are unique when dyadic performance is independently assessed, rather than aggregated.

Discussion

Within this investigation, performance measures were used to describe dyadic sport interactions from three aspects of the same interdependent performance. The purpose was to provide a framework of measurement tools for an applied dyadic performance setting. Competitive cheerleading performances were used to demonstrate the relationships between paired athletes as individual and combined performing units. Significant relationships were observed among the base, flyer, and dyad scores as hypothesized. The individual-level performance scores were strongly correlated, even as products of independent observer pairs. In addition, both the flyer and base scores were moderately associated with the dyad-level score. The key findings support general theoretical interpretations of close partnerships to be relative for the measurement of interdependent sport behaviors.

Individuals' behaviors are determined, in part, by other members of a group (Wageman, 2001). Even when performance can be distinctly distributed among individuals, actions are constrained by the simultaneous and subsequent actions of other team members (Tesluk et al., 1997; Wageman, 2001). Therefore, any measure assigned to a particular athlete will result in performance indicators that take from or add to indications of performance for another athlete (McGarry, 2009). The lack of a theoretical framework forces qualitative task analysis to be partially subjective (Lees, 2002). Determining individual components within the bidirectional influences of team behaviors make performance measurement less than transparent.

Partners' outcomes are interconnected because their behaviors occur within the same performance task (Laursen, 2005; Malloy & Albright, 2001). A heightened similarity exists between partners when compared to any other person in the sample. As a consequence, correlations between individual-level performances are likely to be naturally inflated (Kenny et al., 2006). The strong likeness to one-another causes intact dyad members to typically violate the assumption of independent observations (Gaudreau et al., 2010; Kenny et al., 2006). Even when individual performances can be clearly evaluated, a higher-order effect is still evident (Arthur et al., 2005; Tesluk et al., 1997). As in the current investigation, dyad-level analyses should be considered within the measurement of individuals' interdependent behaviors (Gaudreau et al., 2010).

Interdependence between partners does not automatically eliminate the importance of individual contributions to a relationship (Laursen & Bukowski, 1997). Higher-level performance scores can often result in some information being lost or misinterpreted (Malloy & Albright, 2001). For example, it is typically assumed that each member's input is an equal contribution to the team-level outcome (Tesluk et al., 1997). This assumption is not always the reality. Neglecting individual performance information will result in an incomplete and deficient analysis of a team (Arthur et al., 2005). The current investigation exemplifies individual performance assessments indicative of unequal inputs. The flyer is largely dependent on her partner's performance; suggesting, if any role were more determinant of a dyad-level observation, it would likely be the base. However, the flyers had a slightly stronger correlation, relative to their partners, with the observed dyad scores. Perhaps, the flyer role, in large constraint of her partner's ability, is more telling of a pair's performance. Equal significance of individuals' behaviors cannot always be assumed. Straightforward assessments of lower-level units provide contextual meaning to performance behaviors (McGarry, 2009; Travassos et al., 2013).

Aggregation is commonly used to acknowledge differing levels of the same variable because this technique does not violate statistical assumptions (Kashy, Campbell, & Harris, 2006; Tesluk et al., 1997). However, higher-level constructs represented by aggregated individual data often waste useful information and provide inadequately equal representations of team performance (Malloy & Albright, 2001). Individual scores are not directly indicative of relationship behaviors (Arthur et al., 2005). Within this investigation, the individual and aggregated dyad scores were practically identical performance descriptions. The results reflect the often small benefit gained from collective behavior described by aggregation scores (McGarry, 2009). Aggregation scores are not necessarily useless measures of performance, but should be guided by a strong theoretical rationale, evidence of individual's empirical likeness, and recognition of changing measurement properties (Tesluk et al., 1997).

In line with Wickwire et al.'s (2004) qualitative assessment of intact dyads, this investigation demonstrated the importance of analyzing performance from a multilevel framework highlighting both individual and conjoint contributions. Data from the current investigation was at the descriptive level and continued efforts to critically assess multilevel processes embedded in overarching team outcomes are encouraged (Travassos et al., 2013). Several influences likely exist within performing dyads including within each level (i.e., nonindependence; Kenny et al, 2006), cross-level moderations, and top-down effects present within the dyad-individual hierarchy (Gaudreau et al., 2010). These aspects will largely vary across sports and tasks to the degree to which interdependence dictates athlete interactions and the outcome calls for collective action (Wageman, 2001). Future research featuring causal influences within dyadic sport interactions are encouraged as a more robust test of theories and models surrounding interdependence (Gaudreau et al., 2010).

A particularly important extension from dyadic research entails the study of interdependent actions within differing larger group sizes. McGarry (2009) suggests that player-player dyads offer a basic unit of analysis for investigating space-time dynamics in more traditional team sports and argues the individual within a complex system is centered on dyadic interactions. Current performance analysis approaches that focus on discrete actions in isolation from a meaningful performance context, including team members' actions, have major weaknesses (Travassos et al., 2013). The use of dyadic, subgroup, and team performance analyses in combination offers a more complete picture of coordinated sport behaviors (Travassos et al., 2013).

While researchers have recently considered the emergence of sport behaviors in context of athletes behaving simultaneously, there is a lack of meaningful information that functions to support coaches, athletes, and sport governing bodies (McGarry, 2009; Travassos et al., 2013). Research conclusions shaped for applied sport settings are vital because noncontextual conclusions of sport performance offer little operational advantages (Travassos et al., 2013). Systematic performance analysis of individual contributions and errors within a dyadic interaction may offer solutions for recovering from poor group performances and prevent athletic coordination from deteriorating altogether. The conclusions and procedures within this investigation offer adjustments to the current cheerleading scoring procedures. Suggested adjustments from results in the current investigation include consideration of the multiple levels of coordination present within interactive aesthetic sports. Future research investigating which particular level or combination of levels provides the best representation of performance is needed for better recommendations. Effective scoring systems, guided in scientific principles, can navigate governing bodies to the critical components related to required performance behaviors (McGarry, 2009).

Evidence for reliability of both developed measurement tools indicated modest to satisfactory observer agreement for newly developed measures. Future studies using experienced judges would further indicate the quality of the developed scoring system. Judges can never be trained to the level of perfect agreement because, as human raters, each judge will be associated with errors and inconsistencies (Huang & Foote, 2011; Looney, 2004). Applied performance measurement delicately exists between robust scientific accuracy and the reality of human raters using subjective scales within real time.

The current investigation is limited by the undetected sources of possible measurement biases within subjective observation scores that are difficult to differentiate (Kottner et al., 2011). Often considered as possible sources of measurement error are observers' interpretations of performances and use of different personal standards when applying rating scales (Hoyt & Kerns, 1999). Sex-linked differences, an obvious variable distinguishing the observer pairs for each role in the current investigation, are one example of previously studied influences on impression formation. Specifically, subpar impressions of males' physical attributes, specifically by male observers, have been reported to generate harsher criticism and significantly lower ratings of physical attributes when compared to female observers (Shields, Brawley, & Martin Ginis, 2007). Although bases may have actually been less technically correct than the flyers, role-related differences in performance score distributions may reflect systematic observer biases such as those linked to an observer's sex.

High quality perception requires an observer's brain to be structured and informed towards specific movement patterns for proper interpretation (Knudson, 2013). While observer biases likely were present in the current investigation, strategies were implemented to reduce such effects. Tactics, as reported by Hoyt and Kerns (1999) to minimize a large variance of observer errors, included providing at least five hours of observer training, aggregating scores from multiple observers, using a completely crossed-lagged observation design, and minimizing the occurrence of observer inferences with the rating scales. One limitation within the investigation's design was the lack of a statistically supported reason to terminate observer training. The mentioned protocols may have minimized measurement error related to observer perceptions, but further investigation and would reveal the extent biases and error are represented within each of the performance scores.

Complicated measurement issues surrounding interdependence should not deter researchers from testing intricate relationships in sport (Gaudreau et al., 2010). It was evident that one measure of performance was not a direct indicator of the other two performance perspectives, but in totality, the two-level performance measures provided a conceptually grounded picture of performing dyads. Each partner's individual-level performance score, although related to one another and nested within a dyadic interaction, provided a unique performance indicator relative to the dyad-level performance score. Behavioral exchanges defined by dyadic interactions are vital to team effectiveness and the measurement of the joint connections should be emphasized (McGarry, 2009; Tesluk et al., 1997; Travassos et al., 2013). This investigation generated a foundation in which to study the individual performing within a group, presented adequate confirmation to attend to the multiple levels of performance beyond aggregation when appropriate, and provided a conceptual framework for observing behavioral outcomes in interactive sports.

References

Armstrong, C. W., & Hoffman, S. J. (1979). Effects of teaching experience, knowledge of performer competence, and knowledge of performance outcome on performance error identification. *Research Quarterly for Exercise and Sport 50*(3), 318-327. doi: 10.1080/00345377.1979.10615617

Arthur, W., Edwards, B. D., Bell, S. T., Villado, A. J., & Bennett, W. (2005). Team task analysis: Identifying tasks and jobs that are team based. *Human Factors, 47*(3), 654-669.

Barnett, L., van Beurden, E., Morgan, P. J., Lincoln, D., Zask, A., & Beard, J. (2009). Interrater objectivity for field-based fundamental motor skill assessment. *Research Quarterly for Exercise and Sport, 80*(2), 363-368. doi: 10.1080/02701367.2009.10599571

Biscan, D. V., & Hoffman, S. J. (1976). Movement analysis as a generic ability of physical education teachers and students. *Research Quarterly for Exercise and Sport, 47*(2), 161-163. doi: 10.1080/10671315.1976.10615356

Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment.* Champaign, IL: Human Kinetics.

Franks, I. M., & Goodman, D. (1986). A systematic approach to analysing sports performance. *Journal of Sports Sciences, 4*(1), 49-59. doi: 10.1080/02640418608732098

Fernandez, J., Camerino, O., Anguera, M. T., & Jonsson, G. K. (2009). Identifying and analyzing the construction and effectiveness of offensive plays in basketball by using systematic observation. *Behavior Research Methods, 41*(3), 719-730. doi: 10.3758/BRM.41.3.719

Gangstead, S. K., & Beveridge, S. K. (1984). The implementation and evaluation of a methodological approach to qualitative sport skill analysis instruction. *Journal of Teaching in Physical Education, 3*(2), 60-70.

Gaudreau, P., Fecteau, M.-C., & Perreault, S. (2010). Multi-level modeling of dyadic data in sport sciences: Conceptual, statistical, and practical issues. *Measurement in Physical Education and Exercise Science, 14*(1), 29-50. doi: 10.1080/10913670903455017

Hare, A. P. (1976). *Handbook of small group research* (2nd ed.). New York, NY: Free Press.

Hauw, D., Renault, G., & Durand, M. (2008). How do aerial freestyle skiers land on their feet? A situated analysis of athletes' activity related to new forms of acrobatic performance. *Journal of Science and Medicine in Sport, 11*(5), 481-486. doi: 10.1016/j.jsams.2007.06.005

Hay, J. G., & Reid, J. G. (1982). *The anatomical and mechanical bases of human motion*. Englewood Cliffs, NJ: Prentice Hall.

Hay, J. G., & Reid, J. G. (1988). *Anatomy, mechanics, and human motion*. Englewood Cliffs, NJ: Prentice Hall.

Hoyt, W. T., & Kerns, M. D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*(4), 403-424. doi: 10.1037/1082-989X.4.4.403.

Huang, J., & Foote, C. J. (2011). Using generalizability theory to examine scoring reliability and variability of judging panels in skating competitions. *Journal of Quantitative Analysis in Sports, 7*(3), 1-21.

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sport Sciences, 20*(10), 739-75. doi: 10.1080/026404102320675602

James R. & Dufek, J. S. (1993). Movement observation: What to watch … and why. *Strategies, 7*(2), 17-19. doi:10.1080/08924562.1993.11000276

Kashy, D. A., Campbell, L., & Harris, D. W. (2006). Advances in data analytic approaches for relationships research: The broad utility of hierarchical linear modeling. In A. L. Vangelisti & D. Perlman (Eds.), *The Cambridge handbook of personal relationships* (pp.

73-89). New York, NY: Cambridge University Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83*(1), 126-137. doi: 10.1037//0022-3514.83.1.126

Knudson, D. V. (2013). *Qualitative diagnosis of human movement: Improving performance in sport and exercise* (3rd ed.). Champaign, IL: Human Kinetics.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, Roberts, C., Shoukri, M., Streiner, D. L. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology, 64*(1), 96-106. doi: 10.1016/j.jclinepi.2010.03.002

Laursen, B. (2005). Dyadic and group perspectives on close relationships. *International Journal of Behavioral Development, 29*(2), 97-100. doi: 10.1080/01650250444000450

Laursen, B., & Bukowski, W. M. (1997). A developmental guide to the organisation of close relationships. *International Journal of Behavioral Development*, *21*(4), 747-770. doi:10.1080/016502597384659

Lees, A. (2002). Technique analysis in sports: a critical review. *Journal of Sport Sciences, 20*(10), 813-828. doi: 10.1080/026404102320675657

Levine, J. M., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., pp. 415-469). New York, NY: McGraw-Hill.

Levine, J. M., & Moreland, R. L. (2006). Small groups: An overview. In J. M. Levine, & R. L. Moreland (Eds.), *Small Groups: Key Readings* (pp. 1-10). New York, NY: Psychology Press.

Looney, M. A. (2004). Evaluating judge performance in sport. *Journal of Applied Measurement, 5*(1), 31-47.

Malloy, T. E., & Albright, L. (2001). Multiple and single interaction dyadic research designs: Conceptual and analytic issues. *Basic & Applied Social Psychology, 23*(1), 1-19.

McGarry, T. (2009). Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport, 9*(1), 128-140.

McPherson, M. (1990). A systematic approach to skill analysis. *Science Periodical on Research Technology in Sport, 11*(1), 1-10.

Morrison, C. (2000). Why don't you analyze the way I analyze? *Journal of Physical Education, Recreation & Dance, 71*(1), 22-25. doi: 10.1080/07303084.2000.10605980

National Cheerleading Association (2013). NCA/NDA Competition rule book for college teams. Daytona Beach, FL: National Cheerleading Association.

O'Donoghue, P. (2010). *Research methods for sports performance analysis*. New York, NY: Routledge.

Shields, C. A., Brawley, L. R., & Martin Ginis, K. A. (2007). Interactive effects of exercise status and observer gender on the impressions formed of men. *Sex Roles, 56*(3-4), 231-237. doi: 10.1007/s11199-006-9166-0

Tesluk, P., Mathieu, J. E., Zaccaro, S. J., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 197-224). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Travassos, B., Davids, K., Araújo, D., & Esteves, P. T. (2013). Performance analysis in team sports: Advances from an ecological dynamics approach. *International Journal of Performance Analysis in Sports, 13*(1), 83-95.

Wageman, R. (2001). The meaning of interdependence. In M. E. Turner (Ed.), *Groups at work: Theory and research* (pp. 197-218). New York, NY: Lawrence Erlbaum.

Weekley, J. A., & Geir, J. A. (1989). Ceilings in the reliability and validity of performance ratings: The case of expert raters. *Academy of Management Journal, 32*(1), 213-222. doi: 10.2307/256428

Wickwire, T. L., Bloom, G. A., & Loughead, T. M. (2004). The environment, structure, and interaction process of elite same-sex dyadic sport teams. *The Sport Psychologist, 18*(4), 381-396.

Zelaznik, H. (2014). Errors. In R. C. Eklund, & G. Tenenbaum (Eds.), *Encyclopedia of sport and exercise psychology* (pp. 94-98). Thousand Oaks, CA: Sage Publications. Inc.

Endnotes

[1] A *base* is defined as “any person who is in direct contact with the performing surface and is supporting another person’s weight” and a *flyer* is defined as “any person who is either being supported by another while off of the performing surface or who has been tossed into the air by another person” (NCA, 2013, p. 2).

[2] Public access to the performance videos (for the large co-ed division) are available at the following website (http://varsity.com/event/1725/2013_NCA_NDA_College).

[3] Full list of critical features is available from first author upon request.

Table 1.
NCA Categories of Task Difficulty adapted for Dyad-Level Performance Assessments

Score Range	Category	Descriptions
5 - 6	Beginning Stunt Skills	Shoulder stands Extensions Chair sits
6 - 7	Intermediate Stunt Skills	Liberty (with variations) Awesomes *Includes minimal inverting/twisting/unique transitions, mounts, and dismounts*
7 - 8	Intermediate Stunt Skills	Liberty (with variations) Awesomes *Includes strong incorporation of inverting/twisting/unique transitions, mounts, and dismounts*
8 - 9	Advanced Stunt Skills	Toss one arm and/or one leg stunts to an extended position *Includes strong incorporation of inverting/twisting transitions, mounts, and dismounts*
9 - 10	Elite Stunt Skills	Twisting/inverting mounts into one leg and/or one arm stunts that *also* include inverting/twisting dismounts

1

Table 2.
NCA Categories of Deductions Adapted for Dyad-Level Performance Assessments

Value of Deduction	Category	Descriptions of Errors
- 0.5	Bobbles	Stunts that almost drop/fall, but are saved.
		Incomplete twisting cradles.
		Memory mistakes involving obvious execution of incorrect moves.
		Knee or hand touching ground during cradle or dismount.
		Severe balance checks.
		Severe timing issues.
- 1.0	Mistakes	Drops from stunt that land in a cradle.
		Drops from stunt to a pop down dismount (early dismount).
- 1.5	Falls	Drop to the ground.

Note. There is a maximum deduction of 2.0 points per dyad.