*Journal of Teaching in Physical Education,* 2008, 27, 220-240


# The Game Performance Assessment Instrument (GPAI): Some Concerns and Solutions for Further Development

**Daniel Memmert**
University of Heidelberg

**Steve Harvey**
Leeds Metropolitan University

The purpose of this article is to discuss some concerns with the Game Performance Assessment Instrument (GPAI). This review of the GPAI includes five perceived problems with the GPAI scoring and coding system: (1) calculation of individual and overall game performance indices, (2) use of game involvement versus game performance index to analyze game performance, (3) observer reliability, (4) linear ■unequalness, and (5) usefulness of action. In this article, we suggest a reexamination of the GPAI scoring and coding system that will lead to more efficient use of this game performance instrument. Some of the suggested modifications can be implemented quickly, whereas others need further research.

***Keywords:*** assessment of game play, team sport, invasion games, training, reliability

Traditionally, games lessons are structured around the learning of specific skills. However, alternative teaching approaches have been brought forward in the literature, such as the teaching games for understanding (TGfU) approach (Bunker & Thorpe, 1982), the tactical games model (TGM; Mitchell, Oslin, and Griffin, 2006), play practice (Launder, 2001), and the tactical-decision learning model (Gréhaigne, Wallian, & Godbout, 2005). All these approaches suggest a reversal of the skills-first approach to teaching games, in which teaching is driven by the need for contextual, real-world, game-simulated practice to develop game knowledge and understanding (i.e., knowing what to do, and when and how to do it).

With the evolution of these teaching approaches and the increased focus on teaching the tactical dimensions of game play (Griffin, Oslin, & Mitchell, 1995; Mitchell, Griffin, & Oslin, 1995; Mitchell, Oslin, & Griffin, 1995), there is a

Memmert is with the Institute for Sport and Sport Sciences, University of Heidelberg, Heidelberg, Germany, and Harvey is with the ■Department of ______________, Leeds Metropolitan University, Leeds, U.K.

need for sound, authentic (i.e., genuine and real world; Griffin & Richard, 2003) assessment tools that can (a) discriminate more- and less-advanced levels of game playing ability (Almond, 1986; French & Thomas, 1987; Turner & Martinek, 1992) and (b) help teachers assess what has been taught (Gréhaigne, Richard, & Griffin, 2005; ■Richard, Godbout, & Griffin, 2002). Owing to the difficulty in measuring game knowledge and understanding, it is not surprising that the measurement of game performance is one of the widely examined phenomena in physical education in recent years, and will remain one in future decades (Chen & Rovegno, 2000; Gréhaigne, Godbout, & Bouthier, 1997; Griffin & Richard, 2003; Gréhaigne et al., 2005; Nevett, Rovegno, Barbiaz, & McCaughtry, 2001; Oslin et al., 1998; Richard et al., 2000; ■Richard, Godbout & Griffin, 2002; Richard, Godbout, Tousignant, & Gréhaigne, 1999).

A number of previous studies have attempted to measure game ability using game play protocols (cf. French et al., 1996; French & Thomas, 1987; Jones & Farrow, 1999; McPherson & French, 1991; Turner, 1996; Turner & Martinek, 1992, 1999). In these approaches, decision making and skill execution were categorized to gather independent game play components for each subject (appropriate decisions were coded 1 and inappropriate 0). Gréhaigne et al. (1997) devised the Team Sport Assessment Procedure (TSAP) to measure invasion game play performance, and this instrument has also been adapted to measure game performance in volleyball (Griffin & Richard, 2003; ■Richard, Godbout, & Griffin, 2002). The TSAP is based on the two basic notions of (a) receiving the ball and (b) playing the ball. As with the game play protocols, the TSAP has concentrated on assessing offensive on-the-ball elements of game play performance and involvement.

Oslin, Mitchell, and Griffin (1998) suggested their Game Performance Assessment Instrument (GPAI) to evaluate game performance. The GPAI was intrinsically linked to their development of the aforementioned TGM so that teachers could link what was being taught and learned to the assessment of their students (Griffin & Richard, 2003). The wide acceptance of the GPAI is demonstrated by a number of studies that have used the GPAI for recording data during game play (e.g., Griffin et al., 1995; Harvey, 2003; Mitchell et al, 1995a, 1995b; ■Mitchell & Oslin, 1999). However, although these studies have used the GPAI, there is a lack of follow-up studies that report further validation of the GPAI, especially in the areas of off-the-ball play (both offensive and defensive). Indeed, in using the GPAI in both research and practice, we have found a number of perceived limitations. Thus, after a short introduction of the GPAI, we will report these limitations and offer possible solutions. The suggested modifications to the GPAI will hopefully lead to a more sensitive game performance instrument.

## The Game Performance Assessment Instrument

The Game Performance Assessment Instrument (Oslin et al., 1998) was developed to measure "game performance behaviors that demonstrate tactical understanding, as well as the player's ability to solve tactical problems by selecting and applying appropriate skills" (p. 231). To characterize game play performance in invasion games, it is necessary to identify nonspecific observable components of game performance (Memmert, 2004, 2005; Mitchell et al., 2006). These are crucial in various games, such as soccer, basketball, softball, rugby, or field hockey. To measure single components of game performance, Mitchell et al. (2006)—together with other experts (i.e., trainers, teachers) with knowledge across all four game categories (invasion, net/wall, striking and fielding, and target)—indicated seven tactical components (base, adjust, decision made, skill execution, support, cover, guard/mark) associated with effective game performance (see Table 1 below for nonspecific descriptions of these elements of game play).

Depending on the game and game category, teachers, coaches, and/or researchers can select one or several of the elements of game play seen in Table 1 to evaluate the performance of individuals in that game. Indeed, two benefits of using the GPAI to assess performance are that (a) it can be adapted to various sports and game activities and, (b) it has the ability to not only measure on-the-ball skills, but also off-the-ball skills (both offensive and defensive; Mitchell et al., 2006). For example, in an invasion game, a teacher/coach/researcher may choose to assess on-the-ball components of play, such as *skill execution and decision making*, as well as off-the-ball components, such as how well a player *supports* teammates when their team has possession of the ball. Similarly, when a team does not have the ball, a teacher/coach/researcher can assess how well a player *adjusts* to the changing position of the ball as it is circled by the opposition's offense or how well the player *covers* to provide defensive help to teammates or *guards/marks* players from the opposing team.

Mitchell et al. (2006) detail two methods of scoring game performance with the GPAI: a tally method and a 1–5 Likert-like method. Using Table 2, a teacher could create a set of *1–5 descriptors* for a game, such as rugby, and then assess students based on these descriptors, scoring them 1–5. The original idea of the 1–5 method was to provide an easier way to score performance in faster invasion-type

**Table 1 Game Components Observed in the GPAI (Generic Definitions)**

| Game component | Description |
|---|---|
| Decision making | Makes appropriate decisions about what to do with the ball (or projectile) during a game |
| Skill execution | Efficient Execution of selected skills |
| Adjust | Movement of the performer, either offensively, as necessitated by the flow of the game |
| Cover | Provides appropriate defensive cover, help, backup for a player making a challenge for the ball (or projectile) |
| Support | Provides appropriate support for a teammate with the ball (or projectile) by being in a position to receive a pass |
| Guard/Mark | Appropriate guarding/marking of an opponent who may or may not have the ball (or projectile) |
| Base | Appropriate return of the performer to a recovery (base) position between skill attempts |

games. The second rationale was to allow for a wider range of scoring, but not so wide that reliability was difficult to achieve (Griffin & Richard, 2003). A further benefit of the 1–5 system is its user friendliness, especially for practitioners. The teacher creates a set of 1–5 indicators, which are based on their learning objectives, student abilities, and the school context (see Table 2).

In contrast, the *tally scoring method* was reserved for slower games, such as striking and fielding games (where there was a lot of action followed by a natural break while teams set up for the next pitch), and/or when researchers scored game performance components from videotapes (Mitchell et al., 2006). Indeed, with advancements in technology, and the ability to use computerized software equipment to record performance behaviors, this now enables researchers and practitioners to code game performance components more effectively using the tally system. Tally systems can also be used in peer assessment procedures, like those that have been used with the TSAP (Gréhaigne et al., 2005). It is this tally system that this paper concentrates on.

**Table 2 Behavioral Definitions for Off-the-Ball Support in Rugby**

| Rating/Definition |
|---|
| 5: Very effective performance: |
| *Always* attempting to get open for passes; communicates and demands ball from teammates. Regularly uses sharp cuts to get into open spaces on the field *being involved regularly* in missed pass, switch pass, overlaps, and ball collection after a teammate has been tagged. |
| 4: Effective performance: |
| *Most of the time* tries to get open for passes; communicates and demands ball from teammates. Uses sharp cuts to get into open spaces on the field being involved in *some* moves such as missed pass, switch pass, overlaps, and ball collection after a teammate has been tagged. |
| 3: Moderately effective: |
| Player is beginning to communicate with and demand ball from teammates. Player attempts to get open for passes although cuts to get into open spaces are *slower* and player is only *sporadically* involved in moves and ball collection after a teammate has been tagged. |
| 2: Weak performance: |
| Player *rarely* communicates with and demands ball from teammates. Player attempts to get open to receive passes although cuts to get into open spaces are *slower,* and if the player does not receive the ball *gives up*. Player is *rarely* involved in moves and in ball collection after a teammate has been tagged. |
| 1: Very weak performance: |
| Player *never* communicates with and demands ball from teammates. Player *never* tries to get open to receive passes from teammates and player has *no concept of moves,* such as missed pass, switch pass, overlaps, and *never* collects ball after a teammate has been tagged. |

## Validation

Some of the GPAI components—such as decisions made, skill execution, support, adjust, and game performance (GP)—have previously been validated in the games of soccer, basketball, and volleyball (Oslin et al., 1998). (a) The observer reliability was calculated with the event-recording method. Forty-eight percent of the interobserver agreement (IOA) measures were very high (<.90■), 48% were high (.80–.90), and only once (2%) was below the conventional level of acceptance (<.80). (b) To determine the reliability of the GPAI, the test–retest method was used. For all GPAI components, the stability-reliability coefficients are over the conventional level of acceptance (>.80). (c) The validity of the GPAI was achieved through face validity, content validity, and construct validity. According to the construct validity, in 66% of the cases the results of the GPAI components can distinguished between students ranked high or low in game play by their teachers (Oslin et al., 1998). Although this may seem to be only a moderate value of validity, it is, arguably, more difficult to measure some aspects of game ability than motor skill performance. Alternatively, someone could conclude that some psychometrical problems can arise regarding observation and calculation by mean of the GPAI.

## Coding and Calculating Game Performance Indices

At this point in the article, it therefore seems pertinent to give an example of how the tally coding system works as it is currently constructed (see Mitchell et al., 2006), in order that some of the issues and solutions detailed later in this article can be better understood.

When coding using the tally system, each coder has the responsibility to code behavior each time a player conducts this behavior. These individual observable behaviors are assessed as appropriate/effective or inappropriate/ineffective responses. When the amounts of appropriate/effective and inappropriate/ineffective actions have been totaled, an individual component index can be constructed. For example, we may make one index for decision making and one for skill execution (see Table 3).

In either case, each player starts with a score of 0, and gains 1 point per appropriate decision/effective skill execution and 1 per inappropriate/ineffective (see Mitchell et al., 2006, for more information). For example, to use the decision making column only, our hypothetical player John (see Table 3) makes 20 appropriate decisions and no inappropriate decisions. These two scores are formulated into a decision making index, DMI: $[A/(A + IA) + E/(E + IE)]/2$. In this example, John would score $20/(20 + 0) = 1$. For skill execution, John did not effectively execute any of his appropriate decisions; thus, for the skill execution index (SEI) would score $0/(0 + 20) = 0$. In this way, a player's score always ranges from 0 to 1, and, as suggested by Mitchell et al. (2006), this score can be multiplied by 100 to additionally reveal a percentage. In Table 3, this score is 100% for DMI and 0% for the SEI. In sum, a lower score (i.e., nearer to 0, or 0%) would therefore reflect a player who had more areas of (needed) improvement than a player who scored closer to 1, or 100%.

Two overall indices of performance—game performance (GP) and game involvement (GI)—can be calculated from the aforementioned scores on decision

**Table 3 GPAI Assessment Using Tallies for John**

| | Decision making | | Skill execution | |
|---|---|---|---|---|
| **Name** | **A** | **IA** | **E** | **IE** |
| John's raw score | 20 | 0 | 0 | 20 |
| Indices (%) | DMI = A/(A + IA)<br>DMI = 20/(20 + 0) = 1.0 (100%) | | SEI = A/(A + IA)<br>SEI = 0/(0 + 20) = 0.00 (0%) | |
| John's GP (%) | GP = DMI + SEI/2<br>GP = (1 + 0)/2 = 0.5 (50%) | | | |
| John's GI | GI = sum of all appropriate and inappropriate behaviors (i.e., appropriate decisions made (20) + inappropriate decisions made (0) + appropriate skill executions (0) + inappropriate skill executions (20)<br>GI = 20 + 0 + 0 + 20 = 40 | | | |

*Note.* A = appropriate, IA = inappropriate, E = effective, IE = ineffective, GP = game performance, GI = game involvement.

making and skill execution (see Table 3) to analyze performance improvements and assess overall involvement in the game. The GP index is calculated by summing the two individual indices together and diving by the actual number of indices used (in our example, there are two: DMI and SEI). The GI index is the sum of all the behaviors, so we simply sum the numbers of appropriate/effective and inappropriate/ineffective actions: GI = 2 + 0 + 0 + 2 = 4. ■However, when coding the "off-the-ball" actions, such as adjust, or cover support. GPAI does not count inappropriate actions as involvement, only appropriate.

## Five Concerns With the GPAI

In the next section, we describe what we believe to be five limitations of the GPAI scoring and coding system when using the tally method, and we suggest a solution to each of these problems. The limitations of the GPAI are headed as follows: 1—Calculation of the Individual and Overall Game Performance Indices, 2—Use of Game Involvement Index Versus Game Performance Index to Analyze Game Performance, 3—Observer Reliability, —Linear ■Unequalness, and 5—Usefulness of Action. We have ranked the problems on a scale of "easy" to "hard" and start by discussing Problem 1, the easiest.

### Problem 1—Calculation of Individual and Overall Game Performance Indices

In the way in which the calculations of the index scores are currently used (see Mitchell et al., 2006, and previous section of this article), a player must register in the appropriate/effective category to calculate "true" individual game performance indices (e.g., DMI and SEI). If a player fails to do this (although this may be rare), then the calculation of the index becomes mathematically impossible because zero is the numerator in the index equation and zero is not a divisible number (Rudin, 1987). Using the example of hypothetical player Jill in Table 4, although she makes two inappropriate decisions, the fact that she does not make any appropriate decisions gives her a zero score on the DMI (as 0/(0 + 2) = 0). The same result would occur for a player who made 10 inappropriate decisions (0/(0 + 10) = 0). Our proposition is that there should be some way of calculating this inappropriate decision making within the actual individual game performance index (i.e., decision making, skill execution), rather than relying on the overall GI index to pick this up. Thus, not only would this give a more accurate reflection of the student's/player's score for decision making, but also the teacher/coach/researcher would not have to refer to the overall GI index to pick out the negative involvement. Furthermore, it may be quite de-motivating for a student/player who is putting in a lot of effort to continually score zero on an individual index, even though he or she has been involved, although negatively. Thus, from the way the GPAI index calculations are presently set up, one hopes to get a player who registers in both appropriate/effective and inappropriate/ineffective columns in order to register a score greater than zero. Indeed, in the present method, even if the player made one appropriate decision and two inappropriate decisions, then the DMI would increase from 0 to 0.33, as 1/(1 + 2) = 0.33. Ultimately, as we discuss later, this also affects the calculation of the overall GP index.

***Solution 1.*** Two solutions to this problem are offered. First, one way would be to start each player with a score of 10 (a constant) rather than 1. If player Jodie (see Table 5) did not make any appropriate decisions, but made two inappropriate decisions, and started with a score of 10 and gained 1 point per decision (appropriate or inappropriate), this would give her a score of 10 in the appropriate column and 12 in the inappropriate column. Because we can divide 10 (as 10 is a divisible numerator) by 12, this would allow us to have an index for Jodie even though she did not make any appropriate decisions. Thus, in this case, Jodie's DMI score would be 0.45, or 45%: 10/(10 + 12) = 0.4. We can also calculate the SEI when a player makes no effective skill executions and two ineffective executions (see Table 5). "True" index scores could still be calculated for a player who registered in both columns (see Jason's scores in Table 6). The GP and GI indices (see Tables 5 and 6) would then be calculated using the original method, as suggested by Mitchell et al.

**Table 4 GPAI Assessment Using Tallies for Jill**

| | Decision making | | Skill execution | |
|---|---|---|---|---|
| **Name** | **A** | **IA** | **E** | **IE** |
| Jill's Raw Score | 0 | 2 | 2 | 0 |
| Indices (%) | DMI = 0/(0 + 2) = 0 (0%) | | SEI = 2/ (2 + 0) = 1 (100%) | |
| Jill's GP (%) | GP = (0 + 1)/2 = 1/2 = 0.50 (50%) | | | |
| Jill's GI | GI = 0 + 2 + 2 + 0 = 4 | | | |

*Note.* A = appropriate, IA = inappropriate, E = effective, IE = ineffective, GP = game performance, GI = game involvement.

(2006). In addition, multiplying index scores by 100 (i.e., $0.50 \times 100 = 50\%$) would also reveal a percentage, as suggested by Mitchell et al. (2006). The calculations of percentages are good because they offer a simple way to interpret the index scores for teachers, students, and researchers. It also makes comparisons across different classes (possibly different ages) and population groups easier.

At this point, it may be pertinent to share what a score of 0.45, or 45%, may mean in terms of an "index" score. If a player scored an equal number of decisions in each category, 3 appropriate and 3 inappropriate, then the index score would equal 13/(13 + 13) = 0.50, or 50%. In the example of Jodie (see Table 5), she had more inappropriate actions than appropriate; thus, the index score would be less than 0.50, or 50%. In line with this interpretation of the scores, when a player scored more appropriate actions than appropriate, the index score would be greater than 0.50, or 50%.

Therefore, the higher the index score over 0.50, or 50%, the better the performance because the player makes more appropriate actions, and vice versa, the low performance score (nearer to 0) reflects a weak performer (i.e., one who makes more inappropriate than appropriate actions). There is one issue with this method of calculation in that a performer may not register in either column, but could still score 0.50, unlike in the method suggested by Mitchell et al. (2006) where the performer would score 0. In this rare instance, when a student/player fails to

**Table 5 GPAI Assessment Using Tallies for Jodie**

| | Decision making | | Skill execution | |
|---|---|---|---|---|
| **Name** | **A** | **IA** | **E** | **IE** |
| Jodie's raw score | 0 (+ 10) = 10 | 2 (+ 10) = 12 | 0 (+ 10) = 10 | 2 (+ 10) = 12 |
| Indices | DMI = 10/(10 + 12) = 0.45 (45%) | | SEI = 10/(10 + 12) = 0.45 (45%) | |
| Jodie's GP | GP = (0.45 + 0.45)/2 = 0.9/2 = 0.45 (45%) | | | |
| Jodie's GI | GI = 0 + 2 + 0 + 2 = 4 | | | |

*Note*. A – appropriate, IA – inappropriate, E – effective, IE – ineffective, GP = game performance, GI = game involvement.

**Table 6 GPAI Assessment Using Tallies for Jason**

| | Decision making | | Skill execution | |
|---|---|---|---|---|
| **Name** | **A** | **IA** | **E** | **IE** |
| Jason's raw score | 20 (+ 10) = 30 | 3 (+ 10) = 13 | 8 (+ 10) = 18 | 15 (+ 10) = 25 |
| Indices | DMI = 30/(30 + 13) = 0.70 (70%) | | SEI = 18/(18 + 25) = 0.42 (42%) | |
| Jason's GP | GP = (0.70 + 0.42)/2 = 1.12/2 = 0.56 (56%) | | | |
| Jason's GI | GI = 20 + 3 + 8 + 15 = 46 | | | |

*Note.* A = appropriate, IA = inappropriate, E = effective, IE = ineffective, GP = game performance, GI = game involvement.

register in either column, it is suggested that the teacher/coach/researcher not use the new formula, and scores a student 0 for that index. Thus, the formula is used only when a student/player registers in one of the two columns, appropriate and/or inappropriate. This would, once again, mean that the teacher/coach/researcher would not have to refer back to the GI score to tease this out (see Problem 2).

Secondly, the scores for appropriate and inappropriate actions can also be kept separate. Researchers will report only the individual appropriate/effective and inappropriate/ineffective action scores. The rationale for this is that using an index or ratio may sometimes hide the true nature of where potential strengths and areas for improvement for a player are and, thus, where a teacher/coach can focus the learning/training. For example, if certain players are making high amounts of inappropriate actions, then this player can have practice focused on limiting the number of inappropriate actions. This could be achieved in soccer using an attacking overload play practice where five attackers would play against four defenders. This, with effective coaching, would help the defenders in stressing their ability in limiting their inappropriate actions when under pressure. Indeed, the best indicator of performance and/or improvements in performance can then be completed by examining and reporting the separated individual game performance component scores for appropriate/effective and inappropriate/ineffective actions or by using the involvement index. This point is further examined in Problem 2 below.

## Problem 2—Use of Game Involvement index Versus Game Performance Index

Not only may players not register in a column (appropriate and/or inappropriate, as discussed in Problem 1) and, thus, score zero for that individual game component index (e.g., decision making), but there is also a problem with using indices or ratios because they can mask the true nature of the player's performance profile.

This example is highlighted in Tables 7, 8, and 9. Please note that in the following examples we will use the original index calculations as suggested by Mitchell et al. (2006) and not those suggested in Problem 1 of this article. There are two players: Player A and Player B (see Table 7). Player A may have a DMI and SEI of 0.5 owing to the fact that this player made one appropriate/effective action and one inappropriate/ineffective action for each of the two individual game components (i.e., decisions made and skill executions). However, Player B may have made five appropriate/effective and five inappropriate/ineffective actions and still have a ratio of 0.5 for DMI and SEI. This story holds when the player's overall GP index is examined (see Table 7). Unless we can go back to the individual total scores for each component (e.g., decision making, skill execution), we would never know. Thus, using an index or ratio for either individual component or overall GP gives a false reading of a player's game performance and involvement unless we use their GP index score alongside that of GI. Because GI is the sum of all the behaviors, for Player A, it is 4, and for Player B, it is 20. Now there is an indication as to which player is more involved as well as which player has the better performance.

The same problem can occur for students/players who had high and/or low numbers of appropriate/effective and inappropriate/ineffective actions for each individual game component (see Table 8). In this example, we see the same effect as in our first example where Players A and B have the same GP score (25%; see

**Table 7 Comparison of Two Players' GPs Who Have Different GI Scores**

| | Player A | | | | Player B | | | |
|---|---|---|---|---|---|---|---|---|
| **Component** | **Decision making** | | **Skill execution** | | **Decision making** | | **Skill execution** | |
| A/IA or I/IE | A | IA | E | IE | A | IA | E | IE |
| Index Raw Score | 1 | 1 | 1 | 1 | 5 | 5 | 5 | 5 |
| DMI/SEI | DMI = 1/(1+1) = 0.5 (50%) | | SEI = 1/(1+1) = 0.5 (50%) | | DMI = 5/(5+5) = 0.5(50%) | | SEI = 5/(5+5) = 0.5 (50%) | |
| GP | GP = (0.5 + 0.5) / 2 = 0.5 (50%) | | | | GP = (0.5 + 0.5) / 2 = 0.5 (50%) | | | |
| GI | GI = 1 + 1 + 1 + 1 = 4 | | | | GI = 5 + 5 + 5 + 5 = 20 | | | |

*Note.* A = appropriate, IA = inappropriate, E = effective, IE = ineffective, DMI = decision making index, SEI = skill execution index, GP = game performance, GI = game involvement.

**Table 8 Comparison of Two Players' GPs Who Have Different GI Scores**

| | Player A | | | | Player B | | | |
|---|---|---|---|---|---|---|---|---|
| **Component** | **Decision making** | | **Skill execution** | | **Decision making** | | **Skill execution** | |
| A/IA or I/IE | A | IA | E | IE | A | IA | E | IE |
| Index raw score | 1 | 3 | 1 | 3 | 4 | 12 | 4 | 12 |
| DMI/SEI | DMI = 1/(1 + 3) = 0.25 (25%) | | SEI = 1/(1 + 3) = 0.25 (25%) | | DMI = 4/(4 + 12) = 0.25 (25%) | | SEI = 4/(4 + 12) = 0.25 (25%) | |
| GP | GP = (0.25 + 0.25)/2 = 0.25 (25%) | | | | GP = (0.25 + 0.25)/2 = 0.25 (25%) | | | |
| GI | GI = 1 + 3 + 1 + 3 = 8 | | | | GI = 4 + 12 + 4 + 12 = 32 | | | |

*Note.* A = appropriate, IA = inappropriate, E = effective, IE = ineffective, DMI = decision making index, SEI = skill execution index, GP = game performance, GI = game involvement.

**Table 9 Comparison of Two Players' GP Who Have the Same GI Scores**

| | Player A | | | | Player B | | | |
|---|---|---|---|---|---|---|---|---|
| **Component** | **Decision making** | | **Skill execution** | | **Decision making** | | **Skill execution** | |
| A/IA or I/IE | A | IA | E | IE | A | IA | E | IE |
| Index raw score | 12 | 13 | 10 | 15 | 10 | 15 | 7 | 18 |
| DMI/SEI | DMI = 12/(12 + 13) = 0.48 (48%) | | SEI = 10/(10 + 15) = 0.40 (40%) | | DMI = 10/(10 + 15) = 0.40 (40%) | | SEI = 7/(7 + 18) = 0.28 (28%) | |
| GP | GP = (0.48 + 0.40) / 2 = 0.44 (44%) | | | | GP = (0.40 + 0.28) / 2 = 0.34 (34%) | | | |
| GI | GI = 12 + 13 + 10 + 15 = 50 | | | | GI = 10 + 15 + 7 + 18 = 50 | | | |

*Note.* A = appropriate, IA = inappropriate, E = effective, IE = ineffective, DMI = decision making index, SEI = skill execution index, GP = game performance, GI = game involvement.

Tables 7 and 8) even though Player B's GI is four times that of Player A (see Table 8). This problem is solved only when both players have similar total amount of actions for each of the game components, i.e., GI (see example in Table 9). Table 9 shows that Players A and B can have the same GI (50), but can have a different GP (see Table 9). Thus, GP is truly comparable only when GI is the same. Furthermore, relying on using the GP index on its own, without referring to the GI index, may give a false reading of the number of interactions by the individual player during the game.

***Solution 2.*** Two solutions are offered. First, it may be a good idea to separate the appropriate and inappropriate actions and not make a ratio or index at all (this has already been discussed as a solution to Problem 1).

The second solution is more difficult to resolve. One idea would be to use one index score and not the other (i.e., choose either the GI or the GP). The answer as to which overall index to use may further depend on the level of the class or groups that are being taught, and on the age, experience, and expertise of the players. In a physical education setting, higher involvement may be preferable in the lower grade levels (i.e., grades 4–6). However, at the older grade levels the students are in a position to better understand what needs to be done to better contribute to improved performance. Mistakes by students offer opportune "teachable moments" for the instructor, and Launder (2001) argued that making mistakes comes before, or as a consequence of, success and that only by making mistakes can players learn to be better next time. This is obviously related to opportunities to respond. If such opportunities increase, the hope is that ultimately performance will increase as a consequence of the feedback this provides. This process may also be supported by progressive and skilful instruction (Hopper, 2002). These would include the use of questioning, freeze replays (Launder, 2001), debate-of-ideas settings (Gréhaigne, Godbout, & Bouthier, 2001), or team talks. Thus, GI may be better with the lower grade levels as the students learn to play the game, but GP would be a better indicator of learning and performance improvement with older students, as they look to make more appropriate choices as their game understanding develops.

A second idea would be to translate the overall GP and GI index scores into one overall index (i.e., GI/GP = overall score) to make more effective conclusions about the player's level of game performance and involvement. This is similar to the solution posited in Solution 1, where the inappropriate actions are included in the calculation of the individual index scores. In this way, the amount of inappropriate actions would then be automatically included in the overall GP index calculation. A score of less than 50% would be reflective of a player with more inappropriate actions, and a score over 50% would reflect a player with more appropriate actions. Therefore, GI, in a sense, may become redundant.

A similar solution is offered in the TSAP (Gréhaigne et al., 1997), where both the "volume of play" (VP; like GI) and the "efficiency index" (EI; like GP) are used to create an overall "performance score" (PS). Although this seems like a suitable solution, there are still some issues that remain. First, the pedagogical implications of using such an index need consideration. In the TSAP, Gréhaigne and colleagues consider the VP to be the conquered and received balls by the player, and when calculating the PS divides the VP by 2. Is it appropriate to do this with GI in the GPAI and/or count only the appropriate actions and ignore inappropriate involvement? The EI is the ratio of VP to the amount of balls lost. In the calculation of the PS, the EI score is multiplied by 10, so that scores are on the same scale to allow for easy interpretation. The PS is therefore a combination of the total offensive on-the-ball involvement and the efficiency of these actions.

One further complication in creating one overall PS in the GPAI is that, depending on the pedagogical emphasis placed on the assessment, at present, the GPAI has the ability to additionally consider off-the-ball game involvement in the computation of both the GI and GP indices. In the computation of GI, whereas both appropriate and inappropriate on-the-ball behaviors are considered as involvement, inappropriate off-the-ball movements are not. However, both appropriate and inappropriate on- and off-the-ball behaviors are considered in the computation of the final GP index scores. Therefore, it may be pertinent to give a PS score for both on- and off-the-ball play. These problems could, once again, be dealt with by including *all* inappropriate actions in the calculation of the individual index (i.e., decision making) and overall GP scores. The GI may become redundant because the "involvement" (both positive and negative) has been included in the calculation of overall GP, and a GP of less than 50% would be reflective of a player with more inappropriate actions, and a score over 50% would reflect a player with more appropriate actions. However, the pedagogical implications of such a system need consideration because, even when using this method, greater inappropriate involvement by a student/player would result in a lower individual or overall GP index score. Thus, the GI index may need to be retained for this purpose. Consideration of the different groups being taught would then, once again, become a factor.

Whatever happens, this issue does remain unresolved here. However, further discussion on how to create and compute one overall index score is needed, and would be welcomed, for GPAI, to make more effective use of this instrument in both research and teaching.

## Problem 3—Observer Reliability

In its basic version, the GPAI does not take the results of all the observers into consideration, if there is more than one. The results of all observers are only needed when calculating the observer reliability, and not when considering the actual scoring of behavior. For example, Coder A will score the decision making of player Susan with 3 and player Thomas with 5. This is in contrast to Coder B, who will assess the decision making of Susan with 5 and Thomas with 3. If only Coder B is trusted, then Susan is better than Thomas. If both assessments are averaged, then both players seem to be equal.

***Solution 3.*** For practical implications this is not a big problem because the teacher is normally on his or her own and, thus, is the only person assessing the performance of the children. In this way, the teacher is the expert. However, in research, having the opinion of the assessments of some observers is a better and fairer process. Especially in invasion games, it is often quite difficult to say what an appropriate or inappropriate response is (see Problem 5). The traditional GPAI does not take into consideration the number of observers for each single judgment (appropriate or inappropriate action) of the observers. More exactly, the assessments of the observers will not be averaged because they are isolated summaries of the individual decisions for each observer, respectively. Only now the average assessment as a formula from appropriate and inappropriate actions (see Table 3) of all observers is calculated. Another calculation index is suggested for each game performance component that is observed (Equation 1). To understand the mathematical symbols in this formula please see Hart (2001)

$$GP = 2 \bullet \frac{\sum_{k=1}^{n} (a_a + 1)}{\sum_{k=1}^{n} (a_a + a_i + 2)} \qquad (1)$$

This formula takes into consideration the assessment of all the coders ($k = 1$ to $n$) for appropriate actions ($a_a$) and inappropriate actions ($a_i$), and creates values from 0 to 2 for each coder. If two coders only score the decision making as one component of the GP, this formula means that the assessments of both ($k = 2$) are taken into consideration for the GP of Player 1 simultaneously. Coder A will score the appropriate decision making actions ($a_a$) with 5 and the inappropriate ($a_i$) with 1. Coder B assesses the appropriate decision making ($a_a$) with 4 and the inappropriate ($a_i$) with 3. Using the Equation 1, the GP (= 1.29) will be calculated by 2 × [(5 + 1) + (4 + 1)/(5 + 1 + 2) + (4 + 3 + 2)]. In addition, all actions of each player in every chosen component will be taken into consideration. Moreover, appropriate and inappropriate actions are treated equally (see Problem 4). The traditional GPAI analysis will have calculated .70 by the ■quotation [(5/6) + (4/7)]/2.

This index has another positive feature. Because it creates values between 0 and 2, the results are no longer relative to each other, but are now absolute. In this way, the assessments can be compared with different kinds of groups and other studies if they used the same calculation. All results above 1 indicate that the player is successful and has shown more appropriate than inappropriate actions. In this calculation index GI has to be excluded. As we mentioned regarding Problem 2, we prefer to discuss two kinds of values from each subject separately: GP and GI.

By using the aforementioned method, the problem that different observers could assess different kinds of action is not solved. For example, Coder A evaluated the first decision as inappropriate and the next two as appropriate, and Coder B assessed the first and third decisions of the player as appropriate and the second as inappropriate. Both measurements of decision making will end in the score 2, but with different kinds of assessments per decision. The problem can only be handled if the point in time of each action is registered by the coders as well. This is a problem, and potential solutions are discussed in Problem 5.

## Problem 4—Linear ■Unequalness

The bottom part of the equation of the overall GP is not linear (cf. Hart, 2001; Rudin, 1987). In this way, the good and bad situations are valued differently, that is, not equally strong. Bäumler (2002) gave the example that if decisions made (6 appropriate / 3 inappropriate) and skill execution (3 effective / 6 ineffective) are observed and totaled, there are 9 responses, respectively. In this way, the index can be calculated as 2 for decision made (6/3 = 2) and the index as 0.5 for skill execution ($3/6 = 0.5$). To explain the linear ■unequalness, the GP is calculated by taking the mean out of both results, $[(2 + 0.5)/2 = 1.25]$. We can now see that good responses are treated in a different way than bad ones. Is a good decision valued more than a bad decision and vice versa? Would this depend on whether the ball is moved forward? Therefore, should appropriate responses be weighted differently to inappropriate responses?

***Solution 4.*** This is a difficult problem to solve, and one that will probably remain unresolved in this article. However, we do give a suggestion on how this could be achieved in the solution of Problem 3. In Equation 1, appropriate and inappropriate actions are treated equally. If the numbers from the example in the preceding paragraph are used in this formula—decisions made: 6 appropriate / 3 inappropriate, and skill execution: 3 effective/ 6 ineffective—this yields $2 \times (7/11) = 2 \times [(6 + 1)/(6 + 3 + 2)]$ for decisions made, and $2 \times (4/11) = 2 \times [(3 + 1)/(3 + 6 + 2)]$ for skill execution. In this way, the index can be calculated for decisions made, $14/11 = 2 \times (7/11)$, and for skill execution, $8/11 = 2 \times 4/11$. If the GP was calculated now by taking the mean out of both results, $[(8/11 + 14/11){:}2 = 1.00]$, it can be seen that appropriate responses are treated the same way as the inappropriate ones.

## Problem 5—Usefulness of Action

A final problem of the GPAI is that it is quite hard to realize which action is appropriate or inappropriate, especially when coding the off-the-ball game performance behaviors. Moreover, the mistakes of each observer cannot be taken out of the final calculation. For example, one observer could potentially code 25 appropriate actions and another observer only 5. In other words, different coders may have differing views in categorizing the events (see Problem 4). Thus, this is where the definition of the behaviors being observed becomes important. The subjectivity of certain dimensions of the GPAI makes it difficult to use in classroom settings, especially in a peer assessment context. Teachers need to be aware of this in order to make informed decisions about what dimensions of the instrument to use based on the outcomes they are pursuing. Indeed, van der Mars (1989) has previously stated that behaviors need to be clearly defined. Mitchell et al. (2006) also stated that criteria for observation must be specific and observable. Thus, the test used by the researcher should reflect their specific need, and a research criterion that is narrowly focused and specific may lend itself better to effective game performance assessment. In addition, coders, if used, should receive extensive training in order that they can effectively identify appropriate and inappropriate actions before the actual data coding commences.

***Solution 5.*** Narrower definitions of constructs may be needed to aid coders in specifically and reliably identifying appropriate and inappropriate actions over 80% of the time. An example of a narrowly focused behavioral definition of support in soccer could be as follows:

> Support: Player is in or is moving into space to become available to receive a pass in response to the changing actions of the ball carrier. This movement is at an appropriate angle and distance from the ball carrier. This movement may be accompanied by a call/gesture they want the ball.

For teachers, rubrics of definitions (seen previously in Table 2), can be used and adapted for specific need to evaluate game performance in physical education settings (see also Mitchell et al., 2006, pp. 514–515). Although these definitions are simple, and narrowly focused, whether they would suit all developmental levels needs further investigation.

In another example, Memmert (2004, 2005) investigated the number and unidimensional nature of nonspecific tactical problems in invasion games. A hypothetical path diagram was established to evaluate the statistics. The hypotheses regarding the structure of a data set were tested with a confirmatory factor analysis. To ensure the statistical validation of the path diagram constructed, the maximum likelihood approach was used together with a cross-sectional design to review 21 basic tactical performance parameters from 95 children, within the internal structure of an analysis of moment structures (AMOS; Arbuckle, 1997) model (see Table 10). Seven tactical problems were validated factorially using confirmatory factor analysis ($\chi^2 = 247$; $df = 168$; $\chi^2/df = 1.472$; RMSEA = .071; CFI = .98; AIC = 415; Bollen, 1989). The squared multiple correlation coefficients of the manifest variables are between .20 and .91 for five out of the seven factors. For only two game tactics, the variances of two out of the six indicator variables appear to be too low (<.20). With the exception of one correlation between the latent variables, all parameter estimates are, in part, significantly below the critical value of .90. Only the connection between "support and orienting" and "achieving advantage," which has a value of .97, approaches 1. In accordance with a recommendation made by Arbuckle (1997), one of the parameters should be removed from the model structure because the parameters concerned measure almost identical properties, thereby rendering one of the two redundant. The game components observed in the GPAI decision making category can be observed with these narrower definitions of playing together, using gaps, and fainting.

**Table 10 Overview of the Seven Nonspecific Tactical Problems in Teams' Sports (Memmert, 2004, 2005) Identified by AMOS Analyses**

| Label | Characteristic | Authors |
|---|---|---|
| Attacking the goal | Attacking the goal | Griffin, Mitchell, & Oslin (2006) |
| Taking ball near goal | Setting up to attack | Griffin, Mitchell, & Oslin (2006) |
| | Penetrate—getting close to the goal | Werner (1989) |
| Playing together | Maintaining possession of the ball | Griffin, Mitchell, & Oslin (2006) |
| | Attacking as a team, give and go | Werner (1989) |
| Using gaps | Identification of gaps | Roth (2004) |
| Feinting | Feinting | Werner (1989) |
| Achieving advantage | Create offensive advantage | Werner (1989) |
| | Move without ball | Werner (1989) |
| Support and orienting | Move to open space, using space in attack | Griffin, Mitchell, & Oslin (2006) |

# General Discussion

All the problems we highlight with GPAI, and the solutions, have research and pedagogical implications that need addressing, meaning further discourse on these issues are inevitable. However, some solutions will be of interest for researchers and some solutions could be valuable for teachers in physical education and/or coaching settings.

## Implications for Using the GPAI in Research

The implications of using GPAI in research are that researchers might look carefully into increasing learning time on the GPAI, both before study data are coded, during the coding of study data, and on conclusion of the data-coding process. This will result in more stringent levels of observer reliability being observed through more systematic checks on observer reliability, increasing procedural reliability, and limiting observer drift. That means at least two independent, well-trained coders are needed during coding, and the assessments of all coders have to be taken into consideration (see Equation 1 and Problem 3). In this way, no linear ■unequalness will occur (see Problem 4). In addition, narrower and specific definitions will also help to improve the interrater reliability (see Problem 5). In fact, there has been no validation for a number of constructs in the GPAI; most of these are off-the-ball concepts such as guard/mark, cover, adjust, or base. Indeed, in the original validation levels of observer agreement were much lower for support—an off-the-ball measure that is, arguably, prone to higher levels of intercoder variation owing to the nature of the construct.

We call for one standardized assessment system to assess learning in game play settings. At the moment, the most widely used systems are the GPAI and the TSAP. Both have their advantages. The TSAP focuses more on the offensive on-the-ball aspects of game play, thus, students' involvement on-the-ball, whereas the GPAI has the ability to measure off-the-ball movements. The GPAI could be simplified to take into consideration the two notions of volume and efficiency of play, like the TSAP, and then have one overall GP score, using the reformed calculation system suggested herein (i.e., a "constant" of 10; see Solution 1).

## Pedagogical Implications for Teaching and Learning

The fact that we use authentic assessment, and link what is taught to what is assessed, aids in regulating the teaching and learning process. Gréhaigne et al. (2005, pp. 97–99) suggested the pedagogical implications of using assessment are focused on two primary considerations: (a) the planning cycle, and how individual lessons fit into the yearly curriculum, and (b) the construction of knowledge and skills. Assessment can first help the teacher in the planning cycle, to make the teaching and learning cycle more congruent. For example, the teacher can first confront the students with a problem-solving task (defending space in the court in badminton); students would then play a game (action); and, finally, the students would reflect on their success at the initial goal. Students can then be placed into a situated practice and/or game to help them focus on this initial tactical problem (i.e., a game where they have to return to a "base" position between skill attempts). Use of the GPAI will help focus the students' attention on this particular tactical problem.

This formative assessment process will help students construct knowledge and develop skills, as well as encourage reflective thinking on behalf of the students. They would be able to reflect on areas of strength and improvement via the assessment process, and be able to identify how they can improve. For example, to get back to base more efficiently they may have to hit the shuttlecock into spaces on the opposite side of the court, rather than just getting it over the net. They may also learn that playing the shuttlecock at into these spaces (at varying speeds and heights) will give themselves more time to return to their base position, and be ready for the next shot. The quality of this play would then be assessed in terms of both their GP index score and their GI. In either case, a greater number of appropriate returns to the base position would not only contribute to higher GI scores, but also to a higher GP index on this construct. A higher GP would indicate greater learning of the concept of base position and the initial tactical problem of defending space in the court. This learning could also be assessed in GPAI by concentrating on assessing only single constructs on the GPAI, or converting the GI and GP scores to one overall index, possibly using the new calculation method suggested in this article. However, consideration of the different groups being taught would be a factor. In this instance, however, both inappropriate and appropriate returns to base would help form the basis for the GP index score through use of a constant of 10, with more inappropriate actions resulting in a score of <50% and more appropriate actions resulting in a score of >50% (see Problems 1 and 2).

Finally, the recognition of appropriate and inappropriate game play behaviors in GPAI remains quite subjective, especially for the off-the-ball aspects of play. Thus, teachers and/or teacher educators must undergo a period of training and familiarization (possibly via suitably trained teacher educators) before they use the GPAI in practical settings. This would allow for improved recognition of the game play behaviors and improve the process of integrating the use of GPAI in the teaching and learning process so that proposed learning outcomes can be met successfully by the students/players (see base example given above).

## Summary

Without doubt, the GPAI is one of the two most accepted game performance instruments in the literature. The purpose of this article was to discuss concerns with the GPAI and identify the five possible problems described. The solutions that we offer to each of these five problems do seem somewhat interrelated; however, these solutions need further intensive research. For now, each solution can be adapted to the needs of the individual investigator using the GPAI until further discourse is forthcoming.

Indeed, the first problem can be solved immediately and easily. The problem with calculation of the individual component indices can be rectified with starting every player with a score of 10 (a constant). This further aids in including components of inappropriate involvement into the calculation of the individual index scores, as well as overall GP. Problem 2 noted that the simultaneous use of

GP and GI scores (or including components of inappropriate involvement into the calculation of the individual index scores, as well as overall GP) can be favorable to aid in addressing the issues with using these individual and overall game performance indices, as there is a possibility the use of an index may mask the true level of performance/involvement of the player. More discourse on how to create and compute one overall index score is needed for GPAI, on order to make more effective use of this instrument in both research and in teaching.

The observer reliability could be solved by empirical investigations using Equation 1, given in Solution 3. This formula takes into consideration the assessment of all the coders for appropriate actions and inappropriate actions. Moreover, appropriate and inappropriate actions are treated equally. In this way, as a first step, Problem 4—linear ■unequalness—could be solved. The problem of the usefulness of action, however, will possibly remain unsolved, but starting points could be the use of narrow, specific behavioral definitions and having test definitions that vary or can be applied across a range of developmental levels. These could be coupled with researchers ensuring that coders follow a course of extensive training with these definitions before coding begins.

## Conclusion

We welcome comments and further suggestions on this article; indeed, we call for a reexamination of the GPAI and argue for extending concepts to help us better understand students' game learning (both tactical and/or technical awareness, i.e., what to do, and when and how to do it). Further validation and testing on the off-the-ball components of the GPAI, such as adjust, cover, guard/mark, base, and support, are also needed. Moreover, decision making could also be separated into more specific observable elements, such as playing together, using gaps, and feinting. However, these constructs would also, alongside those off-the-ball components, still need further validation.

The solutions proposed here may help ensure that the measurement of game performance behaviors, especially when using the GPAI, become more sensitive. Moreover, if some of the issues that have been raised are integrated in the future use of the GPAI, it could aid researchers and practitioners by having one globally recognized system that could be used, enabling the gathering of standardized scores of all children across gender, cultures, countries, or developmental levels.

Improving the measurement of game performance behaviors using the GPAI can only enhance our ability to make firmer conclusions about the effects of interventions that aim to improve both the individual and overall components of game performance that the GPAI comprises. Furthermore, improving aspects of measurement and evaluation will aid teachers in training students to perform peer assessment in classroom settings.

■Richard et al. (2002) also concluded that "assessment should be part of everyday teaching" (p. 18), and, therefore, in physical education settings, simultaneous use of assessment and teaching will only further add to the teaching and learning process. It will aid in the teachers' planning cycle of what to teach, and it will also aid in helping the students' construction of knowledge and skills, increasing their ability to recognize game play actions, and translate this learning into performance settings by making more appropriate choices during game play, while limiting inappropriate play. Additional pedagogical tools to illicit critical thinking and reflection could also be used by teachers in this process (i.e., using questioning, freeze replays, debate-of-ideas settings (Gréhaigne et al., 2001), team talks, and possibly video footage), alongside assessment using GPAI, to develop students' game understanding.

## References


Almond, L. (1986). Reflecting on themes: A games classification. In R. Thorpe, D. Bunker, & L. Almond (Ed.), *Rethinking Games Teaching* (p. 71–72). Loughborough: University of Technology.

Arbuckle, J.L. (1997). *AMOS for Windows. Analysis of moment structures (Version 3.6)*. Chicago: University of Chicago Press.

Bäumler, M. (2002). Spielleistungsmessung mit dem Game Performance Assessment Instrument (GPAI). [Game performance measurement with the Game Performance Assessment Instrument (GPAI)]. Abstracts and research program of the 3rd team ball sport conference of the DVS 2002 in Bremen (p. 31–32). Bremen: University of Bremen.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bunker, D., & Thorpe, R. (1982). A model for the teaching of games in secondary schools. *Bulletin of Physical Education, 18,* 5–8.

Chen, W., & Rovegno, I. (2000). Examination of expert and novice teachers' constructivist-orientated teaching practices using a movement approach to elementary physical education. *Research Quarterly for Exercise and Sport, 71*, 357–372.

French, K.E., & Thomas, J.R. (1987). The relation of knowledge development to children's basketball performance. *Journal of Sport Psychology, 9,* 15–32.

French, K.E., Werner, P.H., Rink, J.E., Taylor, K., & Hussey, K. (1996). The effects of a 3-week unit of tactical, skill or combined tactical and skill instruction on badminton performance of ninth-grade students. *Journal of Teaching in Physical Education, 15,* 418–438.

Gréhaigne, J.F., Richard, J-F., & Griffin, L.L. (2005). *Teaching and learning team sports and games*. New York: RoutledgeFalmer.

Gréhaigne, J.F., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education, 16,* 500–516.

Gréhaigne, J.F., Wallian, N., & Godbout, P. (2005). Tactical-decision learning model and students' practices. *Physical Education & Sport Pedagogy, 10*, 255–269. doi:10.1080/17408980500340869.

Gréhaigne, J.F., Godbout, P., & Bouthier, D. (2001). The teaching and learning of decision making in team sports. *Quest, 53,* 59–76.

Griffin, L.L., Mitchell, S.A., & Oslin, J.L. (2006). *Teaching sport concepts and skills: A tactical games approach*. Champaign, IL: Human Kinetics.

Griffin, L.L., Oslin, J.L., & Mitchell, S.A. (1995). An analysis of two instructional approaches to teaching net games. *Research Quarterly for Exercise and Sport* (Suppl.), A-64.

Griffin, L.L., & Richard, J-F. (2003). Using authentic assessment to improve students' net/wall game play. *Teaching Elementary Physical Education, 3,* 23–27.

Hart, M. (2001). *Guide to Analysis*. New York: Palgrave Macmillan.

Harvey, S. (2003). A study of U19 college soccer player's improvement in game performance using the Game Performance Assessment Instrument. *Conference proceedings of the 2nd International Conference for Sport and Understanding, Melbourne.* Retrieved January 19, 2004, from 2003 Teaching Sport for Understanding Conference Web site: http://www.conferences.unimelb.edu.au/sport/proceedings.htm.

Hopper, T. (2002). Teaching games for understanding: the importance of student emphasis over content emphasis. *Journal of Physical Education*, *Recreation, and Dance, 73*(7), 44–48.

Jones, C., & Farrow, D. (1999). The transfer of strategic knowledge: A test of the games classification curriculum model. *Bulletin of Physical Education, 35*, 103–123.

Launder, A.G. (2001). *Play practice: The games approach to teaching and coaching sports*. Champaign, IL: Human Kinetics.

McPherson, S.L., & French, K.E. (1991). Changes in cognitive strategies and motor skill in tennis. *Journal of Sport and Exercise Psychology, 13,* 26–41.

Memmert, D. (2004). Ein Forschungsprogramm zur Validierung sportspielübergreifender Basistaktiken [A research program concerning the validation of non-specific tactical problems]. *German Journal of Sport Science, 34*, 341–354.

Memmert, D. (2005). Zur Identifizierung von Basistaktiken im Sportspiel [Identification of non-specific tactical problems in invasion games]. *Leipziger Sportwissenschaftliche Beiträge, 46,* 33–50.

Mitchell, S.A., Griffin, L.L., & Oslin, J.L. (1995a). An analysis of two instructional approaches to teaching invasion games. *Research Quarterly for Exercise and Sport, 66,* A-65.

Mitchell, S.A., Oslin, J.L., & Griffin, L.L. (1995b). The effects of two instructional approaches on game performance. *Pedagogy in Practice, 1*, 36–48.

Mitchell, S.A., Oslin, J.L., & Griffin, L.L. (2006). *Teaching sport concepts and skills: A tactical games approach* (2nd ed.). Champaign, IL: Human Kinetics.

Nevett, M., Rovegno, I., Barbiaz, M., & McCaughtry, N. (2001). Changes in basic tactics and motor skills in an invasion-type game after a 12-lesson unit of instruction. *Journal of Teaching in Physical Education, 20,* 352–369.

Oslin, J.L., Mitchell, S.A., & Griffin, L.L. (1998). The game performance assessment instrument (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education, 17,* 231–243.

Richard, J-F., Godbout, P., & Gréhaigne, J.F. (2000). Students' precision and interobserver reliability of performance assessment in team sports. *Research Quarterly for Exercise and Sport, 71*, 85–91.

Richard, J-F., Godbout, P., Tousignant, M., & Gréhaigne, J.F. (1999). The try-out of team sport performance assessment procedure in elementary school and junior high school physical education classes. *Journal of Teaching in Physical Education, 18,* 336–356.

Roth, K. (2004). Ballschool Heidelberg. A new answer to traditional questions (unpublished manuscript). *Available at* http://www.ballschule.de/publications/paper.html.

Rudin, W. (1987). *Real and Complex Analysis*. New York: McGraw-Hill.

Turner, A.P. (1996). Teaching for understanding:Myth or reality? *Journal of Physical Education*, *Recreation, & Dance, 67*(4), 46–48.

Turner, A.P., & Martinek, T.J. (1992). A comparative analysis of two models for teaching games. *International Journal of Physical Education, 29,* 15–31.

Turner, A.P., & Martinek, T.J. (1999). An investigation into teaching games for understanding: effects on skill, knowledge, and game play. *Research Quarterly for Exercise and Sport, 70*, 286–296.

Van der Mars, H. (1989). Observer reliability: Issues and procedures. In P. Darst, D. Zakrajsek, & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (2nd ed., pp. 53–80). Champaign, IL: Human Kinetics.

Werner, P. (1989). Teaching games: A tactical perspective. *Journal of Physical Education*, *Recreation, and Dance, 60*, 97–101.