JOURNAL OF TEACHING IN PHYSICAL EDUCATION, 1998, **17**, 231-243

The Game Performance Assessment Instrument (GPAI): Development and Preliminary Validation

Judith L. Oslin and Stephen A. Mitchell
Kent State University

Linda L. Griffin
University of Massachusetts

The purpose of this article is to report on the development and validation of the Game Performance Assessment Instrument (GPAI). The GPAI is a multidimensional system designed 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. The GPAI provides analyses of individual game performance components (e.g., decisions made, skill execution, and support) and/or overall performance (e.g., game involvement and game performance). The individual game performance components were developed and evaluated by experts to determine validity and reliability. The GPAI protocol was field tested across three categories of games: invasion (soccer and basketball), net/wall (volleyball), and field/run/score (softball). Validity and reliability were examined through three separate studies using middle school physical education specialists and their sixth-grade classes. Findings suggest that the GPAI provides a valid and reliable method for assessing game performance.

Games tend to be the predominant curricular activity in secondary physical education. Though rules and strategies for playing games have changed over the years, methods used to teach and assess games have changed very little. Since technical or skill-based methods have been traditionally used to teach sport, skill tests have been promoted as the primary means of student assessment (e.g., American Alliance for Health, Physical Education, Recreation and Dance [AAHPERD], 1989; Kirchner & Fishburne, 1995; National Association for Sport and Physical Education [NASPE], 1992; Pangrazi & Dauer, 1995). Skill tests are generally used before or after a unit of instruction and are used predominantly as a summative form of evaluation (Veal, 1993). Generally, the skill-test grade, combined with the score of a quiz or examination covering the rules and regulations, make up at least

Judith L. Oslin and Stephen A. Mitchell are with the School of Exercise, Leisure, and Sport at Kent State University, 263 Memorial Gym Annex, Kent, OH 44242. Linda L. Griffin is with the Department of Teacher Education and Curriculum Studies at the University of Massachusetts, 103 Furcolo Hall, Amherst, MA 01003.

some proportion of the students' overall grade. The assumption here is that skill-test scores represent the students' ability to perform skills within the game. What the skill-test score does not represent is the ability of the student to perform skills when and where appropriate. For the most part, students' overall game performance is not included as a means of summative or formative evaluation in physical education (Veal, 1993).

Game play is a dynamic event that requires skill proficiency as well as tactical understanding. Though each game is characterized by specific skills, there are a number of tactical similarities between games. Almond (1986) proposed a categorization of games according to their rules and tactical similarities: (a) invasion, (b) net/wall, (c) field/run/score, and (d) target. In invasion games, for example, a team must invade the opposing team's area of the field or court in order to score. Thus, the tactics related to scoring/offense (e.g., maintaining possession, attacking the goal, and creating space in the attack) and preventing scoring/defense (e.g., defending space, defending against an attack, and defending as a team) are similar across invasion games, such as basketball, hockey, and soccer (Mitchell, 1996).

The "Games for Understanding" concept, proposed by Bunker and Thorpe (1982), has initiated a movement toward tactical approaches for games teaching. Initial studies comparing tactical and technical approaches to games teaching (Lawton, 1989; Rink, French, & Werner, 1991; Turner & Martinek, 1992) were equivocal. Some of the more recent studies (McPherson, 1994; Taylor, Hussey, Werner, Rink, & French, 1993; Turner, 1993) have reported significant improvements in declarative knowledge for those subjects receiving tactical instruction when compared to subjects in other treatment groups.

With some exceptions (French & Thomas, 1987; McPherson, 1994; Turner & Martinek, 1992), previous studies have limited dependent measures to include skill development and/or cognitive development (i.e., declarative and procedural knowledge), with the assumption being that improved skill performance leads to improved game play performance. Turner and Martinek's (1992) observation instrument, modified from French and Thomas (1987), included game performance measures related to "on-the-ball" skills (control and execution) and cognitive decisions, but still found no differences between the technical and tactical approaches to instruction.

Though Turner and Martinek (1992) included game performance measures, such measures, similar to previous studies, were primarily associated with skill development. While skill proficiency is important, it is only one facet of game play. In team sports, a considerable portion of game play occurs off-the-ball. In fact, off-the-ball movements and decisions made by supporting players are essential if a team is to be successful (Mitchell, Griffin, & Oslin, 1994). Furthermore, students or players without a high degree of skill can still play the game if they have a tactical understanding of the game. This is the central premise of Games for Understanding. Therefore, in studies of tactical approaches, game performance measures should not only include on-the-ball skills and decisions but also off-the-ball movements to account for the portion of game play that occurs away from the ball.

This article describes procedures related to validating, field testing, and establishing reliability of a protocol for analyzing game performance that includes decisions made and off-the-ball movements. Mitchell, Griffin, and Oslin (1994,

1995) proposed the use of the Game Performance Assessment Instrument (GPAI) as a means of evaluating game play performance. Assuming that enhancement of game play performance is the primary goal of games teaching, then a means of evaluating essential skills and competencies related to game play is necessary and helps uphold the relationship between measures of the independent and dependent variables (i.e., lesson objectives and measures of student learning).

Development of the GPAI

The GPAI provides teachers and researchers with a means of observing and coding performance behaviors (e.g., making decisions, moving appropriately, and executing skills) that are linked to solving tactical problems. Observable components of game performance were initially developed through consultation with teachers and coaches who had expertise in each of the games categories. The aim was to identify observable components of game performance that were applicable across the four game categories. After identifying seven components, descriptions of each component were formulated and reformulated until consensus was reached by all experts (see the following list). When a discrepancy between experts occurred, the definition was reviewed, discussed, and modified. The experts then independently viewed segments of videotaped game play, measured performance using the new definition, and calculated interobserver agreement. There was only one discrepancy, involving the definition of support. The seven game components agreed to were the following:

1. Base: Appropriate return of performer to a "home" or "recovery" position between skill attempts
2. Adjust: Movement of performer, either offensively or defensively, as required by the flow of the game
3. Decisions made: Making appropriate choices about what to do with the ball (or projectile) during the game
4. Skill execution: Efficient performance of selected skills
5. Support: Off-the-ball movement to a position to receive a pass (or throw)
6. Cover: Defensive support for player making a play on-the-ball, or moving to the ball (or projectile)
7. Guard/mark: Defending an opponent who may or may not have the ball (or projectile).

Though all components relate to game performance, not all may apply to a particular game. For example, all components except "base" are important for successful soccer performance. On the other hand, all components except "guard" are important for field/run/score games, such as softball. The GPAI was designed to be a flexible observation instrument that can be used, either with videotape or "live," to observe the performance of any invasion, net/wall, field/run/score, or target game. Teachers or researchers can choose to observe any or all components related to a particular game, depending on the context of the instructional environment. For example, if within a basketball unit a teacher focuses on shooting (skill execution) and support, then only these two components would be included for evaluation. Simplification of the GPAI is especially useful when students are involved in peer or self-evaluation.

Measures of Game Performance

The GPAI can be used to measure individual components of game performance, as well as overall game involvement and performance. For example, the following could be the GPAI outcome variables for soccer:

1. Game involvement = total appropriate responses + number of efficient skill executions + number of inefficient skill executions + number of inappropriate decisions made
2. Decisions Made Index (DMI) = (number of appropriate decisions made) ÷ (number of inappropriate decisions made)
3. Skill Execution Index (SEI) = (number of efficient skill executions) ÷ (number of inefficient skill executions)
4. Support Index (SI) = (number of appropriate supporting movements) ÷ (number of inappropriate supporting movements)
5. Game performance = (DMI + SEI + SI) ÷ 3

The use of appropriate/efficient and inappropriate/inefficient responses provides a more comprehensive view of the students' overall performance. Furthermore, by including inappropriate decisions and inefficient skill executions, students can be evaluated on their game involvement, even though they may not make appropriate decisions or execute skills efficiently.

When scoring individual components, the number of appropriate/efficient responses is divided by the number of inappropriate/inefficient responses. Thus, any score greater than one indicates that a player performed more appropriate/efficient responses than inappropriate/inefficient responses. The use of this type of scoring system eliminates the possibility of any negative scores. Game performance is the average of the individual components included in the index.

Field Testing of the GPAI

Initial field testing of the GPAI was conducted with undergraduate physical education majors (N = 18) enrolled in the Development and Analysis of Team Sports course, which focused on softball, volleyball, soccer, and basketball. During the softball, soccer, and basketball portions of the course, students were paired and requested to observe a single performer during 10 minutes (or three innings) of game play and assess specified components of game performance. During soccer, for example, students observed "decisions made," "skill execution," and "support," since these components had been the focus of instruction during the course. After training, students reported that all components were observable and codable using a simple event recording (tally) format. Interobserver agreement (IOA) for decisions made, skill execution, and support was calculated by means of the scored-interval method (Cooper, Heron, & Heward, 1987) for each pair of observers. Results ranged from 0.66 to 1.0 for soccer, 0.78 to 1.0 for softball, and 0.56 to 0.86 for basketball. The lower end of the ranges in soccer and basketball represent the scores of just a single pair of observers in each case, possibly a result of inexperience in using the instrument. In all instances, students were using the GPAI for the first time. Average IOA was 0.81, 0.86, and 0.83, respectively for soccer, softball, and basketball. An example of the recording sheet used for field testing the GPAI for soccer is shown in the Appendix.

As the coders observed an individual player performing under game conditions, they used the GPAI to assess the appropriateness or efficiency of student/player responses for each component of game performance. For example, during each episode in which the player being observed was in contact with the ball, their evaluator assessed decisions made in terms of passing and shooting decisions. If the player chose to shoot or pass to an open teammate when the opportunity was available, the response was coded as appropriate. If the player being observed did not pass at an appropriate time or passed to a teammate who was not open, the response was coded as inappropriate. Additionally, each time the observed player contacted the ball, the efficiency or inefficiency of his/her passing, receiving, or shooting skill was evaluated.

Sport-specific criteria for an appropriate or efficient rating, established by the content experts prior to observation, are presented in the Appendix. The partner/evaluators also assessed the degree to which the observed player supported, or failed to support, teammates who were in possession of the ball. If the player attempted to support a teammate with the ball by being in or moving into position to receive a pass, the response was recorded as appropriate. However, if the player did not provide support when the opportunity was available, the response was recorded as inappropriate.

Establishing Validity

Face Validity

Though not generally considered necessary to the establishment of criterion-related validity, face validity "may be of importance in determining its acceptability and reasonableness to those who will be tested" (Messick, 1989, p. 60). According to Anastasi (1988), perceived relevance of a test is essential for motivating students to demonstrate their "full repertoire of skills." If tests, such as many of those used to determine motor skill proficiency, are considered to be irrelevant or meaningless, students are likely to be unwilling participants, which compromises the actual validity of the test. Thus, the test-taker's point of view regarding the value of the test is an essential component for determining face validity. According to Raven (1992), "it is meaningless to attempt to assess a person's abilities except in relation to their valued goals" (p. 90).

To determine face validity of the GPAI, undergraduate physical education majors were provided with a questionnaire to assess the degree to which they thought the test was appropriate and fair. The students ($N = 18$) were enrolled in a sophomore-level class, Development and Analysis of Team Sports, and were tested across three sports: softball, soccer, and basketball, then given a modified version of Wiggins' (1993, p. 246) questionnaire.

Overall, student response was quite favorable. When asked whether they preferred to be assessed in a game situation or in a skill-test format, 95% of the students responded positively to being assessed during game play: "It didn't seem like a test, so there was less pressure," "It seems more fair and helps those students that may not be as skilled," and "I liked it. It let me show the teacher what I could do." A number of students recorded negative comments about their previous skill-test experiences: "Skill tests take too long and seem to take up a lot of time" and "If you don't happen to be on the day of the skill test, your grade suffers."

Content Validity

Content validity was determined through a panel of experts (Kerlinger, 1986). Six physical education teacher/coaches with 10 to 30 years of experience in one or more sports were asked to examine and provide feedback on each of the GPAI components. Terms and definitions were revised until all six experts reached consensus on the components detailed in the list provided in the section "Development of the GPAI" above.

Construct Validity

Construct validity of the GPAI was measured by its success in distinguishing between individuals previously rated as high and low in game performance. In three separate studies (soccer, basketball, and volleyball) that compared the tactical approach to a technical (skill-based) approach to games teaching, the middle school physical education specialists were asked a priori to categorize their students' game play performance as high, medium, or low. Students identified as high and low remained as part of the data set. Students identified as medium were not used, thus creating sufficient distance between highs and lows.

In each study, two classes of sixth-grade students participated in small-sided games during two class periods prior to the start of the unit: 6-versus-6 soccer games (n = 53), 3-versus-3 volleyball games (n = 48), and 2-versus-2 basketball games (n = 46). Five minutes of videotaped pretest game play performance was used in the analyses. GPAI components for soccer and basketball included decisions made, skill execution, and support. Components for volleyball included decisions made, skill execution, and adjust.

Pretest game play data were analyzed to determine differences between students rated as either high or low in game play ability by the teacher. Using independent t tests, each component was analyzed separately to determine if the GPAI could differentiate between high and low ability performers. The Bonferroni adjustment (overall alpha/number of tests) was adopted to control the family-wise Type 1 error rate (Wagoner, 1994), with statistically significant differences accepted at the .01 level. As Table 1 indicates, the GPAI was able to differentiate between students rated as high and low ability performers for the skill-execution component across all three studies and the decisions-made component in the soccer and volleyball studies.

To further investigate the differences between high- and low-ability students, and to determine the meaningfulness of the group-difference tests, effect sizes, using pooled standard deviations, were calculated (Thomas & Nelson, 1996). This provides a more complete indication of the GPAI's ability to differentiate between high and low ability performers. As indicated in Table 1, effect sizes (ESs) ranged from 1.93 to 0.23, with six of nine ESs greater than 1.0. According to Thomas and Nelson (1996), "an ES of .8 or greater is large, an ES around .5 is moderate, and an ES of .2 or less is small" (p. 145). Though an ES may be small, such as the 0.23 ES for the adjust component in the volleyball study, any ES greater than zero can be considered to represent meaningful group differences (Thomas & Nelson, 1996). Limited statistical significance can possibly be explained by the large number of medium ability students eliminated from the analyses, thereby decreasing the degrees of freedom. It should also be noted that the conservative Bonferroni adjustment used

Table 1 Comparison of GPAI Components With Students Ranked High or Low in Game Play Performance: A Test of Construct Validity

	Soccer				Basketball				Volleyball			
	High (n = 10)	Low (n = 22)			High (n = 7)	Low (n = 18)			High (n = 12)	Low (n = 19)		
	M	*M*	*t*	*ES*	*M*	*M*	*t*	*ES*	*M*	*M*	*t*	*ES*
Decisions made	2.92 (3.00)	0.76 (0.94)	2.19*	1.19	3.51 (1.69)	2.54 (1.45)	1.44	0.64	2.33 (1.42)	0.75 (0.74)	3.58**	1.50
Skill execution	3.09 (2.10)	0.52 (0.79)	3.66**	1.93	6.58 (2.78)	3.04 (1.89)	3.69***	1.64	1.50 (1.13)	0.28 (0.42)	3.56**	1.58
Support	5.21 (4.98)	1.40 (1.17)	2.36*	1.31	3.23 (2.70)	2.14 (1.99)	1.09	0.50	—	—	—	—
Adjust	—	—	—	—	—	—	—	—	0.33	0.21	0.65	0.23

Note. Standard deviations are given in parentheses.
*$p < .05$. **$p < .01$. ***$p < .001$.

in these analyses required a .01 level of significance rather than the .05 level required by other methods of analysis, such as commonly used MANOVA/ANOVA combinations.

Ecological Validity

Ecological validity, as used here, relates to the consistency between instructional objectives and measures of student performance (Davis & Burton, 1991). In this context, therefore, "the instrument should reflect what is taught" (Rink, French, & Graham, 1996, p. 501). The components of the GPAI were selected and the criteria defined according to what was to be taught in each particular instructional setting. For example, in the soccer study, support, skill execution (passing, receiving, and shooting on goal), and decision making were the focus of the instructional unit. Continuing this example, definitions of efficient and inefficient skill execution were based specifically on what skill elements were presented by the instructor. Similarly, criteria for support and decision making were based on what was taught. The GPAI components of base, marking/guarding, and cover were not assessed in these studies, because content related to these components was not taught. In short, it would not be ecologically valid to assess components that were not addressed during instruction. Furthermore, action research requires ecologically valid measures if findings are to be of value to all participants.

Reliability

Instrument Reliability

To determine the reliability of the GPAI as an instrument, the test-retest method was used to obtain the stability-reliability coefficient (Baumgartner & Jackson, 1991). A random numbers table was used to select students to retest. Retesting was completed on a minimum of 30% of the students in each of the soccer, basketball, and volleyball studies. Videotapes were reviewed and data was recorded on the GPAI recording sheets. Test-retest correlations were then computed for the GPAI components (see Table 2).

According to Baumgartner and Jackson (1991), the closer the stability-reliability coefficient is to positive one (+1), the more reliable the scores. The soccer, basketball, and volleyball correlation coefficients ranged from .84 to .97, .84 to .99, and .85 to .97, respectively (see Table 2). With all test-retest reliability coeffi-

Table 2 Stability-Reliability Coefficients for GPAI Components

	Decisions made	Skill execution	Support	Adjust
Soccer	.847	.971	.865	—
Basketball	.848	.844	.993	—
Volleyball	.896	.850	—	.972

cients above .80, the decisions made, skill execution, support, and adjust components of the GPAI were considered reliable within the parameters of each study.

Observer Reliability

To date, three studies of the tactical approach have used the GPAI to assess game play performance of sixth-grade soccer, sixth-grade volleyball, and sixth-grade basketball. In each study, two coders were used, both of whom were trained to use the GPAI. Training for the coders continued until they consistently exceeded 80% IOA. Using a minimum of 15% of the data, average IOA in each of the three studies ranged from 73% to 97%, with only one pretest category averaging below the conventional level of acceptance, 80% (Cooper, Heron, & Heward, 1987). Table 3 provides IOA measures for each component of the GPAI used in each study, as well as overall IOA measures.

Discussion

Four components of the GPAI were tested to determine their degree of validity and reliability for assessing game play performance across three games, two from the invasion category (soccer and basketball) and one from the net/wall category (volleyball). All components met the requirements for face, content, and ecological validity. In terms of construct validity, conservative data analyses revealed the skill-execution component to be a valid measure of game performance across all three games. The decisions-made component was also determined to be valid, $t(31) = 3.56$, $p < .001$, in the volleyball study, but in the soccer study was only significant at the .05 level, $t(32) = 2.19$. The support component was also significant at the .05 level in the soccer study, $t(32) = 2.36$. However, all ESs that ranged from 1.93 to 0.23 represent meaningful group differences (Thomas & Nelson, 1996). This suggests that the GPAI was able to differentiate between high and low ability performers for each of the game components studied. All four components were found to be amenable to reliable measurement. These preliminary results are encouraging, though further study and replication across grade levels and various games categories is warranted.

Table 3 Average Interobserver Agreement Measures for Three Studies That Used Components of the GPAI to Assess Pre- and Postgame Play Performance

GPAI components	Soccer		Basketball		Volleyball		Overall
	Pre	Post	Pre	Post	Pre	Post	
Decisions made	0.93	0.91	0.80	0.86	0.94	0.93	0.90
Skill execution	0.73	0.84	0.94	0.97	0.90	0.92	0.88
Support	0.81	0.89	0.95	0.90	—	—	0.89
Adjust	—	—	—	—	0.81	0.89	0.85
Overall	0.81	0.86	0.90	0.91	—	—	0.88

As with any instrument, it is important to control the context in which it is applied. One problem arose as coders observed two low ability teams playing a 2-versus-2 game of basketball. The students were instructed to play a half-court game, but implemented a number of modifications to accommodate their game play abilities. For example, players did not contest inbounds passes, the inbounds pass was always made from the end line rather than center court, and they frequently ignored minor violations. Essentially, these players changed the game considerably, necessitating the exclusion of some data from the analysis. Therefore, it is important to control the game setting when using the GPAI.

Another caution relates to the number of players playing and the selection of the components to be assessed. For example, in a 2-versus-2 game of basketball, a player *must* move to support his or her teammate because there is nobody else to provide support. On the other hand, in a game of 3-versus-3 or greater, any one player has the opportunity *not* to support since there is the possibility that another teammate will fulfill this role. To adequately assess a player's ability to provide support, a 3-versus-3, 4-versus-4, or 5-versus-5 game would likely provide a more authentic or valid context in which to assess this component. The use of a 2-versus-2 game situation in the basketball study may account for the nonsignificant difference between high and low ability students on the support component (see Table 1). Compared to basketball, it is interesting to note the more substantial difference in support in the 6-versus-6 soccer game. A future study is planned that will compare basketball game performance of players on various size teams (e.g., 3-vs.-3 compared to 4-vs.-4).

Though still in its infancy, the GPAI may be viable, not only as a research tool, but as a method of authentic assessment that can be and has been used by a number of physical educators. According to Veal (1993), "performance tests often test students on something that is completely unrelated to successful game play" (p. 95). Measures of performance taken during game play provide a more accurate representation of a player's or student's ability. Furthermore, these measures are more authentic, because they occur within the context of the game. Measures of off-the-ball movement corroborate whether a student fully understands his or her role as a supporting member of a team and need to be included in student evaluations of game performance.

Unlike other instruments (e.g., French & Thomas, 1987; Turner & Martinek, 1992), the GPAI includes components to assess game performance off-the-ball, such as support and adjust. As we have shown elsewhere (Mitchell, Oslin, and Griffin, 1995), "decisions or movements made by players who are not in possession of the ball are critical elements of game performance" (p. 45). Furthermore, players spend the majority of game time performing off-the-ball movements, and far less time performing on-the-ball skills. For example, in baseball, the players not playing the ball provide back-up, base coverage, as well as verbal support for the player fielding the ball. While skill execution is important (e.g., fielding, batting, and baserunning), the support play of teammates who are not in possession of the ball is vital to the overall success of the team. Another advantage of including off-the-ball movements is that, in most instances, less complex movement patterns are required for successful execution. Thus, students with lower levels of skill development can still get involved in game play by putting themselves in a more likely position to receive a pass from a teammate or, defensively, by marking/guarding effectively. In a study on expert–novice differences of youth base-

ball players, French, Spurgeon, and Nevett (1995) reported low-skilled players to be less accurate at positioning than high-skilled players. Potentially, low-skilled players could benefit most from instruction related to appropriate off-the-ball movements.

The GPAI provides a means of thinking more broadly about game performance and assessing game play in its entirety. The components of decisions made, skill execution, support, and adjust were observable and showed adequate validity and reliability, but future studies must field test and determine further the validity and reliability for these and other components of game play (e.g., base, covering, and marking/guarding). Also, future studies should use the GPAI to analyze game performance within different games, both within and across games categories, as well as game performance of students or players representing various developmental levels.

Outcomes related to game performance encompass a broad range of skills, movements, and cognitive decisions. Traditionally, in games research and in games instruction, we have failed to consider alternative outcomes and appropriate measures of these outcomes. Though early in its development, the GPAI appears to provide a promising addition to, or alternative for, skill testing. Furthermore, though skill does contribute to game performance, without tactical awareness and understanding it is difficult, if not impossible, to play a game.

References

Almond L. (1986). Reflecting on themes: A games classification. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 71-72). Loughborough, UK: Loughborough University of Technology.

American Alliance for Health, Physical Education, Recreation and Dance. (1989). *Physical best instructor's guide*. Reston, VA: Author.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Baumgartner, T.A., & Jackson, A.S. (1991). *Measurement for evaluation in physical education and exercise science* (4th ed.). Dubuque, IA: Brown.

Bunker, D., & Thorpe, R. (1982). A model for the teaching of games in secondary schools. *Bulletin of Physical Education*, **18**, 7-10.

Cooper, J.O., Heron, T.E., & Heward, W.L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill.

Davis, W.E., & Burton, A.W. (1991). Ecological task analysis: Translating movement behavior theory into practice. *Adapted Physical Activity Quarterly*, **8**, 154-177.

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

French, K.E., Spurgeon, J.H., & Nevett, M.E. (1995). Expert-novice differences in cognitive and skill execution components of youth baseball performance. *Research Quarterly for Exercise and Sport*, **66**, 194-201.

Kerlinger, F. (1986). *Foundations of Behavioral Research* (2nd ed.). New York: Holt, Rinehart & Winston.

Kirchner, G., & Fishburne, G.J. (1995). *Physical education for elementary school children* (9th ed.). Madison, WI: Brown and Benchmark.

Lawton, J. (1989). Comparison of two teaching methods in games. *Bulletin of Physical Education*, **25**, 35-38.

McPherson, S.L. (1994). The development of sport expertise: Mapping the tactical domain. *Quest*, **46**, 223-240.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational Measurement (3rd ed., pp. 60-68). New York: Macmillan.

Mitchell, S.A. (1996). Teaching invasion games for understanding: Improving game performance. *Journal of Physical Education, Recreation and Dance*, **67**(2), 30-33.

Mitchell, S.A., Griffin, L.L., & Oslin, J.L. (1994). Tactical awareness as developmentally appropriate for the teaching of games in elementary and secondary physical education. *The Physical Educator*, **51**, 21-28.

Mitchell, S.A., Griffin, L.L., & Oslin, J.L. (1995). *A comparison of two instructional approaches to teaching invasion games*. Poster session presented at the annual meeting of the American Alliance for Health, Physical Education, Recreation and Dance, Portland, OR.

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

National Association for Sport and Physical Education. (1992). *Outcomes of quality physical education programs*. Reston, VA: National Association for Sport and Physical Education/American Alliance for Health, Physical Education, Recreation and Dance.

Pangrazi, R.P., & Dauer, V.P. (1995). *Dynamic physical education for elementary school children* (11th ed.). Needham Heights, MA: Allyn & Bacon.

Raven, J. (1992). A model of competence, motivation, and behavior, and a paradigm for assessment. In H. Berlak (Ed.), *Assessing student performance: Exploring the purpose and limits of testing* (pp. 89-90). New York: State University of New York Press.

Rink, J.E., French, K., & Graham, K.C. (1996). Implications for practice and research. *Journal of Teaching in Physical Education*, **4**, 490-502.

Rink, J.E., French, K., & Werner, P. (1991, January). *Tactical awareness as the focus for ninth grade badminton*. Paper presented at the World Congress, Atlanta, Georgia, 1991.

Taylor, J.K., Hussey, K.G., Werner, P.H., Rink, J.E., & French, K.E. (1993). *The effects of strategy, skill, and strategy and skill instruction on skill and knowledge development in ninth grade badminton*. Poster session presented at the annual meeting of the American Alliance for Health, Physical Education, Recreation and Dance, Washington, DC.

Thomas, J.R., & Nelson, J.K. (1996). *Research methods in physical activity* (3rd ed.). Champaign, IL: Human Kinetics.

Turner, A.P. (1993). *A model for working with students with varying knowledge structures*. Paper presented at the annual meeting of the American Alliance for Health, Physical Education, Recreation and Dance, Washington, DC.

Turner, A.P., & Martinek, T.J. (1992). A comparative analysis of two models for teaching games (technique approach and game-centered (tactical focus) approach). *International Journal of Physical Education*, **29**, 15-31.

Veal, M.L. (1993). The role of assessment and evaluation in secondary physical education: A pedagogical view. In J.E. Rink (Ed.), *Critical crossroads: Middle and secondary school physical education* (pp. 93-99). Reston, VA: National Association for Sport and Physical Education.

Wagoner, K.D. (1994). *Descriptive discriminant analysis: A follow-up procedure to a "significant" MANOVA*. Workshop presented at the American Alliance for Health, Physical Education, Recreation and Dance, Denver, CO.

Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.

Appendix

Data Sheet for Game Performance Assessment

GPAI: Soccer Study

Coder: _____ Player: _____

Observation: Pregame/Postgame (Circle)

Category—Criteria for appropriate/efficient rating:

1. Decisions made: Player chooses to pass to an open teammate
   Player chooses to shoot when appropriate
2. Skill execution: Reception—Control of pass and set up of the ball
   Passing—Ball reaches target
   Shooting—Ball stays below head height and is on target
3. Support: The player appeared to attempt to support the ball carrier by being in/moving to an appropriate position to receive a pass

Decision made		Skill execution		Support	
A	IA	E	IE	A	IA

Key: A = appropriate; IA = inappropriate; E = efficient; IE = inefficient.