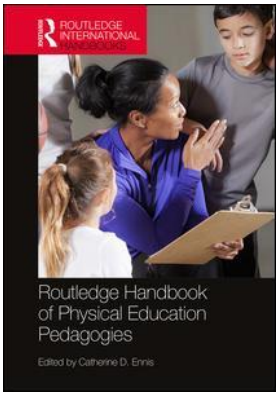


# Routledge Handbook of Physical Education Pedagogies

Catherine D. Ennis

## Measurement of teaching in physical education

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9781315743561.ch26
Prithwi Raj Subramaniam, Deborah A. Wuest
**Published online on: 15 Aug 2016**

# 26

# MEASUREMENT OF TEACHING IN PHYSICAL EDUCATION

*Prithwi Raj Subramaniam & Deborah A. Wuest,* ITHACA COLLEGE, USA

Teaching is a complex multidimensional act involving teachers and students working in partnership to teach and learn from each other (Campbell, Kyriakides, Muijs, & Robinson, 2003; Danielson, 2001; Muijs, Campbell, Kyriakides, & Robinson, 2005; Rink, 2013; Shulman, 1986; Siedentop, 2002; Silverman, 1994). To gain some degree of insight into the complexities involved in the teaching act requires a systematic approach in conducting research and collecting data that are reliable and valid. Quantitative measures that produce valid and reliable scores on student outcome measures are "essential first steps" (Locke, 1977, p. 11) for measuring teaching and enabling researchers to establish causal links between the multiple variables influencing teaching and student learning (Floden, 1986; Rink, 2013). Causal connections between teaching and learning and the interaction of multiple variables in the teaching context can also be established through carefully designed interventions with a clear focus on the desired outcome measure.

How researchers collect data and the type of data collected vis-à-vis outcome measures have a direct impact on the measurement of teaching. Siedentop (2002, p. 427) states that "good data are the key to successful research." Such data permit researchers to provide valid and reliable interpretations and generalizations on effective teaching and learning. Using quantitative measures to gather data allows researchers to gain a deeper understanding of the teaching-learning relationship in making informed decisions to improve teaching practice and student achievement. In an era of standards-based systems of accountability, such data on measurement of teaching become invaluable to teachers and other stakeholders in education. This chapter will acquaint the reader with a brief history of research on measurement of teaching in physical education (PE), quantitative methods of measuring teaching in PE, and suggested directions for future research on teacher effectiveness.

## A brief history of measurement of teaching in physical education

Efforts to measure teaching effectiveness originated in classroom research and have been greatly influenced by the seminal work of Dunkin and Biddle (1974). Dunkin and Biddle advanced a model for the study of teaching that emphasized the interaction between four constructs – presage variables, process variables, product variables, and contextual variables. Early efforts to study teacher effectiveness were based on the process-product paradigm (Shulman, 1986) utilizing direct systematic observation of teacher behaviors, student

behaviors, and the interaction between teachers and students in an effort to understand the relationship between teacher and student variables (Graber, 2001; Lee, 2003; Lee & Solmon, 2005; Rink, 2003). The lack of consistent results, however, coupled with recognition of the complex nature of teaching, the interdependence of student and teacher behaviors, and impact of teacher behaviors on student learning led to a shift to the mediating process paradigm.

In the mediating process paradigm, scholars examined the role of student behaviors as mediators of instruction. Researchers viewed students as active and controlling agents in the learning process (Lee, 2003; Lee & Solmon, 2005; Rink, 2003; Siedentop, 2002). Pedagogy researchers began to use time-based and response-based variables as proxy measures of student achievement using systematic observation (Siedentop, 2002; van der Mars, 2006). The mediating process-product paradigm also extends to classroom ecology (Doyle, 1977) in the study of teaching effectiveness. Using task analysis Doyle emphasized the complex nature of classroom ecology and the influential nature of teacher-student interactions (Hastie & Siedentop, 1999, 2006; Placek & Locke, 1986; Siedentop, 2002; Tousignant & Siedentop, 1983). With the changing conception of teaching and learning, the quest to understand how cognitive processes influence thinking and decision making became an important area of inquiry. Studying student and teacher cognition provided a lens to understand how cognition mediates learning (Lee, 1997, 2003; Lee & Solmon, 2005; Solmon, 2006; see Solmon's Chapter 33 in this volume).

## Quantitative methods of measuring teaching in physical education

Nixon and Locke (1973) created the impetus for systematic observation of teaching by urging PE pedagogy researchers to engage in more descriptive-analytic research. As a corollary to this stimulus, the Teachers College Data Bank Project produced the first descriptive time data on how teachers spent their time in the gymnasium (Anderson & Barrette, 1978). Systematic observation quickly became the *sine qua non* in the measurement of teaching in the field. There was a surge in the development of observation instruments to observe teacher and student behavior through direct observation or via videotape (Placek & Locke, 1986; Silverman, 1991; van der Mars, 2006). The proliferation of direct systematic observation instruments ultimately resulted in the compilation of a compendium containing specialized instruments for use in PE and sport settings (Darst, Zakrajsek, & Mancini, 1989).

### *Systematic observation approach*

Systematic observation is an explicit observation methodology that involves precise specification of what is to be observed, ignored, and recorded (McKenzie & van der Mars, 2015; Michaels, 1983; van der Mars, 1989a). Direct systematic observation involves trained observers to observe, encode, and recode the behavior of participants (Carlson & McKenzie, 1984; McKenzie & van der Mars, 2015; Michaels, 1983). This methodology can provide specific quantitative (i.e., activity type, frequency, and duration) and qualitative (i.e., context) data (Carlson & McKenzie, 1984; McKenzie & van der Mars, 2015; van der Mars, 1989a). A large majority of research on teaching in PE (RT-PE) involving instructional time, sometimes described as opportunity to learn, has depended on direct systematic observation to collect data (Darst et al., 1989; van der Mars, 2006). Typically, researchers using systematic observation use one or more of the four basic observation tactics: event recording, duration recording, interval recording, and momentary time sampling (Carlson & McKenzie, 1984; McKenzie & van der Mars, 2015; van

der Mars, 1989b). These tactics are used to measure/estimate the frequency and/or duration of teacher and student behavior (van der Mars, 1989b, 2006).

## *Event recording*

Event recording is the appropriate tactic for collecting data when the event and/or behavior are repeatable. The repeated occurrence of an event or behavior results in a frequency of that event or behavior. Event recording provides data on the frequency of occurrence of a *discrete event* by tallying the number of times that event occurred. In essence, it provides a frequency count of behavior and the raw data are converted to rate per minute and percent total, and sometimes are reported as ratios (McKenzie & van der Mars, 2015; van der Mars, 1989b). It is important that the behaviors to be observed are clearly defined prior to using this observation tactic to collect data. Van der Mars (1989b) cautions against the use of event recording when a behavior occurs at extremely high rates because keeping count of the extremely high rate of behavior and maintaining reliability could become problematic. Another instance when this tactic is inappropriate is when a behavior or event lasts for an extended period of time. In this case, the frequency data do not convey the duration of the behavior and the validity of the data could be jeopardized (van der Mars, 1989b).

Several studies have used event recording independently or in combination with other systematic observation tactics (e.g., Ashy, Lee, & Landin, 1988; Byra & Coulon, 1994; Lee & Ward, 2009; Rhoades & Woods, 2012; Silverman, 1985, 1988; Silverman, Subramaniam, & Woods, 1998; Solmon & Lee, 1996). An event recording tool currently gaining popularity among RT-PE researchers is the Game Performance Assessment Instrument (GPAI) (Oslin, Mitchell, & Griffin, 1998; van der Mars, 2006) that permits data collection in authentic game play settings.

## *Duration recording*

Duration recording is the appropriate technique to use when systematic observation involves time-dependent variables. This is the preferred technique when the focus is on a temporal dimension of a discrete behavior or event that has a clear beginning and ending (Carlson & McKenzie, 1984; van der Mars, 1989b, 2006). Duration recording is used in measuring continuous behaviors that last for extended periods of time, and the data are typically expressed in minutes or percent of observed time (Carlson & McKenzie, 1984; McKenzie & van der Mars, 2015; van der Mars, 1989b, 2006).

This direct systematic observation is used to collect data on how time is used by students and teachers. The collection of data on time-based variables allows researchers to assess the duration of time a class/student or teacher engages in a particular behavior (Carlson & McKenzie, 1984). Early studies on time-based variables used duration recording (e.g., Anderson & Barrette, 1978; McKenzie, Clark, & McKenzie, 1984). These studies provided the initial conceptualization and systematic understanding of time-dependent variables in the gymnasium (Metzler, 1989). As researchers' conceptualization and understanding of time became more refined, they began to use duration recording in combination with other direct observation tactics (e.g., Cusimano, 1987; Silverman, 1985).

## *Interval recording*

If the intent is to measure the (non)occurrence of a behavior during specified time intervals, then interval recording is the appropriate direct observation method. Typically, interval length

ranges from 3 to 10 seconds (McKenzie & van der Mars, 2015; van der Mars, 2006). Interval recording is a form of discontinuous time-based measurement (Carlson & McKenzie, 1984). Data obtained in interval recording are converted to percentage of intervals (McKenzie & van der Mars, 2015; van der Mars, 1989b, 2006). Observers have the choice of selecting either the *partial-interval recording* or the *whole-interval recording*. In the former procedure, even a fleeting occurrence of the behavior would be coded whereas in the latter procedure the behavior must be present for the whole length of the interval (McKenzie & van der Mars, 2015; van der Mars, 1989b).

The recognition that instructional time could be used as a proxy for student achievement (Siedentop, Tousignant, & Parker, 1982) paved the way for the systematic observation instrument, Academic Learning Time-Physical Education (ALT-PE), to become a widely used means for collecting time-dependent data in PE using the interval recording tactic (Dodds, 1983; Mancini, Wuest, Clark, & Ridosh, 1983; Metzler, 1983; Placek & Randall, 1986; Shute, Dodds, Placek, Rife, & Silverman, 1982; Silverman, Dodds, Placek, Shute, & Rife, 1984; Silverman, Tyson, & Marrs-Morford, 1988). Interval recording can provide estimates of both frequency and duration of behaviors which makes it the most widely used observation tactic (McKenzie & van der Mars, 2015; van der Mars, 1989b, 2006).

## *Momentary time sampling*

The occurrence of individual or group behavior can be measured using momentary time sampling. In this tactic, the coding decision about a behavior occurs at the end of each observational interval (McKenzie & van der Mars, 2015; McNamee & van der Mars, 2005; van der Mars, 1989b). Interval lengths in this direct systematic observation method generally range from 1 to 10 minutes (van der Mars, 1989b). Data are typically expressed as percent of total intervals (individuals) or percent of total number of people observed across all samples (group) (McKenzie & van der Mars, 2015). This tactic has the advantage of permitting observers to use the time in between observations to collect other relevant data (McKenzie & van der Mars, 2015; van der Mars, 1989b). On the other hand, there is the potential for over- or underestimating observed behavior since momentary time sampling only provides an estimate of the occurrence of the behavior (McKenzie & van der Mars, 2015; van der Mars, 1989b, 2006).

System for Observing Fitness Instruction (SOFIT) is the most widely used momentary time sampling instrument in RT-PE. This systematic observation tool has been used successfully in large intervention studies to simultaneously assess physical activity, lesson context, and teacher behavior (McKenzie & van der Mars, 2015). Several studies have used momentary time sampling to study teacher behavior and physical activity in PE (e.g., Chow, McKenzie, & Louie, 2009; Faucette & Patterson, 1990; Kulinna, Silverman, & Keating, 2000; Logan, Robinson, Webster, & Rudisill, 2015; van der Mars, Rowe, Schuldheisz, & Fox, 2004).

## *Advantages and limitations of systematic observation*

There are several advantages of using systematic observation as a methodology in measuring teaching. First, it is a direct method with strong internal (or content) validity that can provide objective data. Second, it permits observers to record behaviors that occur naturally in the teaching context. Third, specific behaviors or events can be targeted to better understand how the teaching environment impacts learning. Fourth, systematic observation provides the lens for researchers to scrutinize the teaching act in a systematic way for analysis, critique, and refinement (Cheffers, 1977). Fifth, the results of descriptive-analytic studies provide the basis

for conducting experimental and evaluative studies (Anderson, 1971). Finally, the data obtained using systematic observation (e.g., frequency, minutes, percent) are easily understood by teachers and administrators (McKenzie & van der Mars, 2015).

Systematic observational methods also have several limitations including observer effects or reactivity. Teachers' awareness of the observer's presence is likely to affect their behavior consciously or unconsciously (Muijs, 2006). For some teachers, awareness of being observed increases the frequency of targeted behaviors and in the process has the potential to influence data fidelity (McKenzie & van der Mars, 2015; Michaels, 1983). Another limitation of systematic observation is the potential for observer bias. Systematic error or bias occurs when human observers are involved no matter how careful the observer is (Michaels, 1983).

Observers' judgmental decisions and inferences during encoding could pose a reliability issue as well. This issue can be minimized through proper training prior to data collection. There also is the potential for observer drift (misinterpretation of defined categories and coding conventions through the passage of time) which warrants recalibration to re-establish reliability (McKenzie & van der Mars, 2015; Michaels, 1983). The variability of a teacher's behavior within and between lessons may be overlooked when data are expressed in percentage or frequency counts (Kyriacou & Newson, 1982).

In addition, using distinct categories of behavior fragments the teaching process by imposing the observer's predetermined selection of important elements of behavior which may not be congruent with the teacher's intentions (Casey & Taylor, 2004; Grant, Ballard, & Glynn, 1986). Moreover, the categories adopted are not sensitive to the quality of the teacher's behavior. For example, observers are not able to discern the quality of questioning when the focus of the coding category is questioning frequency (Kyriacou & Newson, 1982; Michaels, 1983; Muijs, 2006). Additionally, a teacher's behavior may convey meaning to students that cannot be captured through direct observation using systematic observation instruments. Also, the focus on targeted behavior fails to capture the dynamic quality of teacher-student interaction in the learning context (Kyriacou & Newson, 1982; Muijs, 2006). Systematic observation instruments were developed to measure teaching based on direct instruction and may not be a good fit for some model-based instruction. Finally, direct observations of teacher behavior are mere snapshots and do not provide a full picture of teacher behavior in the overall scheme of teaching and learning. They are not intended to measure teacher cognition such as teacher beliefs or subject matter knowledge (Muijs, 2006). Based on the aforementioned limitations of direct systematic observation to analyze teaching effectiveness, caution should be exercised in interpreting such data on teaching effectiveness.

Systematic observation paved the way to studying teaching effectiveness in a more systematic manner in PE and produced the foundational knowledge to our understanding of the teaching-learning process in the gymnasium. There is an extensive body of literature using systematic observation as a data collection method in teacher effectiveness research investigating the relationship between time (and opportunity) and student learning that has generated a descriptive database on time variables in PE (van der Mars, 2006). There also is a growing body of literature on systematic observation tools and physical activity in the school setting. Systematic observation has its advantages and limitations like any other data collection methodology. It will continue to be used appropriately to capture certain types of data in teaching effectiveness research. The traditional paper-and-pencil approach to systematic observation, however, is being supplemented by new technology to facilitate the collection of quality data on teaching effectiveness. The following section will address the use of technology in collecting observational data.

## *Technology and observational data*

Technology has become a powerful tool for collecting observational data in PE. When studying teaching effectiveness, researchers can incorporate appropriate technology to collect and analyze data effectively and efficiently both in terms of teaching behaviors and student outcomes. Researchers' ability to interface computing devices and software has contributed to greater ease in the use of systematic observation tools and an influx of new tools and methodologies for use in research efforts to measure outcome variables (McKenzie & van der Mars, 2015).

Systematic observation users, once tied to paper-and-pencil approaches to data collection, now benefit from portable, compact computer devices – laptops, tablets, and smartphones – that expedite data collection (Castellano, Perea, Alday, & Mendo, 2008). The growing presence of smartphones and tablets and the development of systematic applications ("apps") for the popular iOS (Apple) and Android platforms contribute to greater ease in observation of variables of interest.

Technological advancements in software have facilitated observers' use of various systematic observation tools. Advances cited by McKenzie and van der Mars (2015) include the ability to store videotaped data for later coding, facilitating initial as well as ongoing reliability checks. Stored records make it possible to code videos with other systematic observation instruments, gaining a different perspective on what has transpired. Other advances that hold promise for use in the study of teacher effectiveness are better synchronization of data from multiple sources, including those from systematic observations of teachers and/or students, videos, sensors such as those capturing heart rate or movement, and data from assessments of student outcomes (Zimmerman, Bolhuis, Willemsen, Meyer, & Noldus, 2009). The ability to accurately and reliably synchronize data provides a more complete picture of the teaching-learning process.

Videotaping has long been a staple of movement observation and continues to be an important technique in the assessment of student learning. Advances in video recording and analysis have revolutionized videography. Additionally, the development of mobile devices (tablets and smartphones) and a host of "apps" for analysis and visual representation of performance have facilitated this process. The shift from analog to digital formats has made video easier to use and analyze. The ability to wirelessly transmit video to cloud storage, a computer, or to a mobile device makes the video readily available for analysis. New software and apps (e.g., Coach's Eye, Ubersense, Dartfish Express) create easily understood, visual representations of the data, thus enhancing the ability of teachers and/or their students to analyze their performance, while providing a permanent record for later analysis or sharing.

The miniaturization of sensors with increased storage capacity and rapid data processing along with wireless data transmission and increased affordability offer new possibilities for assessment (Kelly, 2014; McKenzie & van der Mars, 2015; Zimmerman et al., 2009). Wearable sensors, such as pedometers, heart rate monitors, accelerometers, and combined accelerometer and heart rate monitors provide a means of objectively measuring physical activity and its various components. Butte, Ekelund, and Westerterp (2012) discuss various devices, best practices, and future development.

The development of "smart" sports equipment with embedded or attached sensors offers another means to assess sports skill performance, while creating a permanent record and quantifiable data on parameters of interest. Companion apps store information and turn raw data into easily understood displays representing specific performance aspects. Tennis racquets, basketballs, and soccer balls with embedded sensors and sensors to attach to golf clubs and bats now exist, and more "smart" equipment and sensors, along with companion apps, are in development and will likely be available for many popular sports (Kelly, 2014).

Game play is the preferred venue for the authentic assessment of skills and tactical knowledge in many sports. It is challenging to provide a valid and reliable assessment of both players' skills and tactics exhibited during game play. However, current videographic technology coupled with more sophisticated software makes it possible to address this need. Chow, Tan, Lee, and Button (2014) describe how an automated motion-tracking system can be used in PE to authentically assess the tactical dimensions and skill execution during game play. Installation of a video camera directly above the court captures the movements of players. The A-Eye software refines the data, allowing players to be tracked automatically and continuously. Analysis yields information on players' behavior during play, offering insights relative to differences in court coverage and skills execution, such as passing and receiving, or tennis stroke execution. It can also be used effectively to examine players' movements in net and territorial games. The A-Eye provides quantifiable records for teachers to use for student assessment. Capturing information over a number of lessons allows teachers to document students' progress. Chow et al. (2014) note the "relevance and validity of using the A-Eye to examine performance efficiency is strong" (p. 456), "allowing for the more effective, objective and reliable measures of game performance and movement behaviors" (p. 460). Further, it has "good potential in addressing difficulty in the measurement of game performance" (p. 460) in PE.

There is, however, one caution to keep in mind when using technology as a tool for collecting observational data. Technology is like any other systematic observational instrument that requires rigorous procedures (e.g., calibration, training, etc.) for researchers to produce reliable and valid data. The quality of data should not be compromised when using technology to assist in collecting observational data. Technology is subject to measurement error just as are human observers.

Technological advances have provided and will continue to provide greater ease in the use of systematic observation tools in teacher effectiveness research. Advances in software, video recording and analysis, wearable sensors, and automated motion tracking system allow researchers to collect data expeditiously and measure outcome variables objectively. Researchers should practice caution when using technology because technology in itself does not guarantee or improve data accuracy. Using technology in concert with systematic observation instruments has the potential to yield quality data.

## Conclusion and future directions

Researchers use systematic observation to procure reliable and accurate data on teachers' instructional behavior and student outcomes. Systematic observation has provided a rich body of literature on teaching effectiveness. Early studies relied on duration recording to collect data on how teachers and students used instructional time in PE. The acceptance of instructional time as a proxy for student achievement led PE pedagogy researchers to use interval recording for collecting time-dependent data. Event recording is used to collect data when the event and/or behavior are repeatable. A renewed approach to using event recording emerged as a result of a shift from technique-based to game-based approaches to teaching (e.g., GPAI) coupled with authentic assessment practices (Mitchell, Oslin, & Griffin, 2003; Oslin et al., 1998). Momentary time sampling tools (e.g., SOFIT) have been used successfully in large-scale physical activity intervention studies to code behavior at the end of each observational interval. It should be pointed out that these methods have been used independently or in combination with other systematic observation tactics in teacher effectiveness research.

Future teacher effectiveness research needs to focus on developing more ecologically valid instruments that allow researchers to measure student learning in authentic game play settings. Data obtained using these instruments can provide credible evidence of student learning. For example, students not in possession of the ball in territorial games would be coded as not accruing ALT-PE using the interval recoding tactic. However, from a tactical awareness perspective, they are still actively engaged in off-the-ball tactics such as positioning, guarding, and supporting (van der Mars, 2006). While measuring these behaviors in isolation does not fully capture the essence of student learning, using more sophisticated instruments, aided by technology, can assist in enhancing performance assessment.

The broadening conception of teacher accountability and teaching effectiveness (Good, 2014; Kyriakides, Campbell, & Christofidou, 2002) is dependent on accurate ways to measure teacher behavior and student learning. In the future systematic observation instruments will need to be more sensitive to context and instructional models to provide valid and reliable measures. Systematic observation instruments that yield authentic student performance data during game play settings, such as the GPAI (Oslin et al., 1998), can serve this function well. Although using standalone direct observation tactics may not be appropriate to encapsulate the complexity of teaching, combining several direct observation tactics in a single study has the potential to provide rich data on teaching effectiveness. More advanced observation systems can facilitate the design of developmentally appropriate assessment tasks that permit students to demonstrate their learning in authentic game play (van der Mars, 2006).

The changing landscape in games instruction requires measurement tools that are more sensitive in assessing student learning. In this regard, hybrid systematic observation tools that link time/opportunity with newer variables such as decision making and execution will enable researchers to assess student learning in holistic and authentic ways (Rink, 1996; van der Mars, 2006).

Systematic observation tactics use varying coding formats (e.g., interval length, length of observe/record intervals), but there is a paucity of research investigating the variability of coding formats. This is particularly important in ascertaining how far the interval length can be stretched in a particular activity without losing data accuracy (McKenzie & van der Mars, 2015). For example, McNamee and van der Mars (2005) found that interval lengths up to 90 seconds can provide acceptable levels of accuracy for physical activity data in physical education.

There also is a dire need to engage in longitudinal studies to track teaching effectiveness over time in PE. Our counterparts in health have been successful in tracking student behavior changes using longitudinal studies (Summerfield, 2001). The dearth of longitudinal studies in teacher effectiveness research in PE can be attributed to a lack of consensus on outcome measures and the lack of available practical, reliable, and valid measures of student achievement, other than fitness.

Researchers will rely more on technology in the future to collect observational data using smartphones and apps and incorporating automated motion-tracking systems (Chow et al., 2014) to complement and supplement systematic observational data in authentic game play settings. The advent of affordable drones with enhanced capabilities such as wireless transfer of video offers the possibility of more sophisticated data related to game play performance for the entire game, including off-the-ball actions. Possibilities for research expand with the potential to incorporate such technology in conjunction with systematic observation instruments like the GPAI (Oslin et al., 1998).

Future research on teacher effectiveness also will rely on technology-based data collection tools in large-scale multi-site research projects (McKenzie & van der Mars, 2015). A team of researchers will have the capability to access data collected from different sites simultaneously.

Data also can be shared and transferred readily among researchers with expertise in diverse areas allowing for multifaceted analyses. Additionally, systematic observation data collected using technology in both quantitative and qualitative studies can complement and supplement teacher effectiveness research resulting in powerful findings that will help us better understand the complex multidimensional act of teaching.

Progress has been made in measuring teaching effectiveness in PE by shifting our focus from studying individual variables to multiple variables and their influence on teaching and learning using systematic observation tools and technology. Regardless of the systematic observation tactic or technology used to collect data, it is critical that the data produced – whether from observations of teaching or measures of student learning – are valid and reliable.

## Summary of key findings

- Teaching is a complex act. RT-PE has evolved from traditional studies examining individual variables associated with teacher behavior and student outcomes to more contemporary research investigating multiple variables. Systematic observation has been used successfully in the measurement of teaching in efforts to better understand the teaching-learning process.
- Descriptive-analytic research in PE using systematic observation pioneered the creation of a descriptive database on instructional time variables.
- Event recording, duration recording, interval recording, and momentary time sampling are the four basic observation tactics used in systematic observation. Researchers have used them independently or in combination to study the multidimensional act of teaching. A rich body of literature on PE teacher effectiveness research has evolved from systematic observational data.
- Systematic observation is suited to collecting many types of data in teacher effectiveness research. Collecting valid, reliable, and accurate data is the pulse of systematic observation.
- Technological advancements have made major contributions in collecting and analyzing data effectively and efficiently using systematic observational tools.

## Reflective questions for discussion

1. In what ways can systematic observation be used to measure multidimensional elements of teaching? Are there some aspects of teaching that are simply too complex to measure using systematic observation?
2. How can researchers measure the changing roles of teachers and their influence on student learning in teacher effectiveness research?
3. What constraining factors limit researchers' ability to use systematic observation to measure all relevant dimensions in teaching?
4. How and to what extent can Danielson's framework be adapted to systematically observe teachers in PE?
5. In what ways are researchers using technology to study teaching and teacher effectiveness?
6. To what extent does the use of technology-based data collection tools affect data accuracy in teacher effectiveness research?
7. What trade-offs do researchers have to confront in teacher effectiveness research using systematic observation and technology?

## References


Anderson, W. G. (1971). Descriptive-analytic research on teaching. *Quest*, *15*(1), 1–8.

Anderson, W. G., & Barrette, G. T. (Eds.). (1978). What's going on in gym: Descriptive studies of physical education classes. *Motor skills: Theory into practice* [Monograph 1].

Ashy, M. H., Lee, A. M., & Landin, D. K. (1988). Relationship of practice using correct technique to achievement in a motor skill. *Journal of Teaching in Physical Education*, *7*, 115–120.

Butte, N. F., Ekelund, U., & Westerterp, K. R. (2012). Assessing physical activity using wearable monitors: Measures of physical activity. *Medicine and Science in Sports & Exercise*, *44*, S5–S12.

Byra, M., & Coulon, S. C. (1994). The effect of planning on the instructional behaviors of preservice teachers. *Journal of Teaching in Physical Education*, *13*, 123–139.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, *29*, 347–362.

Carlson, B. R., & McKenzie, T. L. (1984). Computer technology for recording, storing, and analyzing temporal data in physical activity settings. *Journal of Teaching in Physical Education*, *4*, 24–29.

Casey, C. E., & Taylor, A. W. (2004). Teacher education research in physical education. *Avante*, *10*(2), 24–31.

Castellano, J., Perea, A., Alday, L., & Mendo, A. H. (2008). The measuring and observation tool in sports. *Behavior Research Methods*, *40*, 898–905.

Cheffers, J. (1977). Observing teaching systematically. *Quest*, *28*(1), 17–28.

Chow, B. C., McKenzie, T. L., & Louie, L. (2009). Physical activity environmental influences during secondary school physical education. *Journal of Teaching in Physical Education*, *28*, 21–37.

Chow, J. Y., Tan, C. W. K., Lee, M. C. Y., & Button, C. (2014). Possibilities and implications of using a motion-tracking system in physical education. *European Physical Education Review*, *20*, 444–446.

Cusimano, B. E. (1987). Effects of self-assessment and goal setting on verbal behaviors of elementary physical education teachers. *Journal of Teaching in Physical Education*, *6*, 166–173.

Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership*, *58*(5), 12–15.

Darst, P. W., Zakrajsek, D. B., & Mancini, V. H. (1989). *Analyzing physical education and sport instruction* (2nd edn). Champaign, IL: Human Kinetics.

Dodds, P. (1983). Relationships between academic learning time and teacher behaviors in a physical education majors skills class. In T. Templin & J. Olson (Eds.), *Teaching in physical education* (pp. 173–184). Champaign, IL: Human Kinetics.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 163–198). Itasca, IL: Peacock.

Dunkin, M., & Biddle, B. (1974). *The study of teaching*. New York, NY: Holt, Rhinehart & Winston.

Faucette, N., & Patterson, P. (1990). Comparing teaching behaviors and student activity levels in classes taught by PE specialists versus nonspecialists. *Journal of Teaching in Physical Education*, *9*, 106–114.

Floden, R. E. (1986). Research on effects of teaching: A continuing model for research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3–16). New York, NY: Macmillan.

Good, T. L. (2014). What do we know about how teachers influence student performance on standardized tests: And why do we know so little about other student outcomes? *Teachers College Record*, *116*(1), 1–41.

Graber, K. (2001). Research on teaching in physical education. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 491–519). Washington, DC: American Educational Research Association.

Grant, B., Ballard, K. D., & Glynn, T. (1986). Directions in research on teaching physical education. *Educational Psychology*, *6*(1), 19–35.

Hastie, P., & Siedentop, D. (1999). An ecological perspective on physical education. *European Physical Education Review*, *5*(1), 9–29.

Hastie, P. A., & Siedentop, D. (2006). The classroom ecology paradigm. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 214–225). Thousand Oaks, CA: Sage.

Kelly, H. (2014, December 31). Smart sports equipment turns phones into coaches. Retrieved from http://edition.cnn.com/2014/11/28/tech/innovation/smart-sports-equipment/.

Kulinna, P. H., Silverman, S., & Keating, S. D. (2000). Relationship between teachers' belief systems and actions toward teaching physical activity and fitness. *Journal of Teaching in Physical Education*, *19*, 206–221.

Kyriacou, C., & Newson, G. (1982). Teacher effectiveness: A consideration of research problems. *Educational Review*, *34*(1), 3–12.

Kyriakides, L., Campbell, R. J., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *13*, 291–325.

Lee, A. M. (1997). Contributions of research on student thinking in physical education. *Journal of Teaching in Physical Education*, *16*, 262–277.

Lee, A. M. (2003). How the field evolved. In S. Silverman & C. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (2nd ed., pp. 9–25). Champaign, IL: Human Kinetics.

Lee, A. M., & Solmon, M. A. (2005). Pedagogy research through the years in RQES. *Research Quarterly for Exercise and Sport*, *76*, S108–S121.

Lee, M., & Ward, P. (2009). Generalization of tactics in tag rugby from practice to games in middle school physical education. *Physical Education and Sport Pedagogy*, *14*, 189–207.

Locke, L. F. (1977). Research on teaching physical education: New hope for a dismal science. *Quest*, *28*, 2–16.

Logan, S. W., Robinson, L. E., Webster, E. K., & Rudisill, M. E. (2015). The influence of instructional climate on time spent in management tasks and physical activity of 2nd grade students during physical education. *European Physical Education Review*, *2*, 195–205.

Mancini, V., Wuest, D., Clark, E., & Ridosh, N. (1983). In T. Templin & J. Olson (Eds.), *Teaching in physical education* (pp. 197–208). Champaign, IL: Human Kinetics.

McKenzie, T. L., Clark, E. K., & McKenzie, R. E. (1984). Instructional strategies: Influence on teacher and student behavior. *Journal of Teaching in Physical Education*, *3*(2), 20–28.

McKenzie, T. L., & van der Mars, H. (2015). Top 10 research questions related to assessing physical activity and its context using systematic observation. *Research Quarterly for Exercise and Sport*, *86*, 13–29.

McNamee, J., & van der Mars, H. (2005). Accuracy of momentary time sampling: A comparison of varying interval lengths using SOFIT. *Journal of Teaching in Physical Education*, *24*, 282–292.

Metzler, M. (1983). Using academic learning time in process-product studies with experimental teaching units. In T. Templin & J. Olson (Eds.), *Teaching in physical education* (pp. 185–196). Champaign, IL: Human Kinetics.

Metzler, M. (1989). A review of research on time in sport pedagogy. *Journal of Teaching in Physical Education*, *8*, 87–103.

Michaels, J. W. (1983). Systematic observation as a measurement strategy. *Sociological Focus*, *16*, 217–226.

Mitchell, S. A., Oslin, J. L., & Griffin, L. L. (2003). *Sport foundations for elementary physical education*. Champaign, IL: Human Kinetics.

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation: An International Journal of Theory and Practice*, *12*, 53–74.

Muijs, D., Campbell, J., Kyriakides, L., & Robinson, W. (2005). Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement*, *16*, 51–70.

Nixon, J. E., & Locke, L. (1973). Research on teaching in physical education. In M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 1210–1242). Chicago: Rand McNally.

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.

Placek, J. H., & Locke, L. F. (1986). Research on teaching physical education: New knowledge and cautious optimism. *Journal of Teacher Education*, *37*(4), 24–28.

Placek, J. H., & Randall, L. (1986). Comparison of academic learning time in physical education: Students of specialists and nonspecialists. *Journal of Teaching in Physical Education*, *5*, 157–165.

Rhoades, J. L., & Woods, A. M. (2012). National board certified physical education teachers task presentations and learning environments. *Journal of Teaching in Physical Education*, *31*, 4–20.

Rink, J. E. (1996). Tactical and skill approaches to teaching sport and games [Monograph]. *Journal of Teaching in Physical Education*, *15*, 397–516.

Rink, J. E. (2003). Effective instruction in physical education. In S. Silverman & C. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (2nd ed., pp. 165–186). Champaign, IL: Human Kinetics.

Rink, J. E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport*, *84*, 407–418.

Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3–36). New York, NY: Macmillan.

Shute, S., Dodds, P., Placek, J., Rife, F., & Silverman, S. (1982). Academic learning time in elementary school movement education: A descriptive analytic study. *Journal of Teaching in Physical Education*, *1*(2), 3–14.

Siedentop, D. (2002). Ecological perspectives in teaching research. *Journal of Teaching in Physical Education*, *21*, 427–440.

Siedentop, D., Tousignant, M., & Parker, M. (1982). *Academic Learning Time-Physical Education Coding Manual: 1982 Revision*. Unpublished manual, The Ohio State University.

Silverman, S. J. (1985). Relationship of engagement and practice trials to student achievement. *Journal of Teaching in Physical Education*, *5*, 13–21.

Silverman, S. J. (1988). Relationship of presage and context variables to achievement. *Research Quarterly for Exercise and Sport*, *59*, 35–41.

Silverman, S. (1991). Research on teaching in physical education. *Research Quarterly for Exercise and Sport*, *62*, 352–364.

Silverman, S. (1994). Research on teaching and student achievement. *Sport Science Review*, *3*(1), 83–90.

Silverman, S., Dodds, P., Placek, J., Shute, S., & Rife, F. (1984). Academic learning time in elementary school physical education (ALT- PE) for student subgroups and instructional activity units. *Research Quarterly for Exercise and Sport*, *55*, 365–370.

Silverman, S. J., Subramaniam, P. R., & Woods, A. M. (1998). Task structures, student practice, and skill in physical education. *Journal of Educational Research*, *91*, 298–306.

Silverman, S., Tyson, L., & Marrs-Morford, L. M. (1988). Relationships of organization, time, and student achievement in physical education. *Teaching and Teacher Education*, *4*, 247–257.

Solmon, M. A. (2006). Learner cognition. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 226–241). Thousand Oaks, CA: Sage.

Solmon, M., & Lee, A. M. (1996). Entry characteristics, practice variables, and cognition: Student mediation of instruction. *Journal of Teaching in Physical Education*, *15*, 136–150.

Summerfield, L. M. (2001). School health education. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 472–490). Washington, DC: American Educational Research Association.

Tousignant, M., & Siedentop, D. (1983). A qualitative analysis of task structures in required secondary physical education classes. *Journal of Teaching in Physical Education*, *3*, 47–57.

van der Mars, H. (1989a). Systematic observation: An introduction. In P. W. Darst & D. B. Zakrajsek (Eds.), *Analyzing physical education and sport instruction* (2nd ed., pp. 3–17). Champaign, IL: Human Kinetics.

van der Mars, H. (1989b). Basic recording tactics. In P. W. Darst & D. B. Zakrajsek (Eds.), *Analyzing physical education and sport instruction* (2nd ed., pp. 19–51). Champaign, IL: Human Kinetics.

van der Mars, H. (2006). Time and learning in physical education. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 191–213). Thousand Oaks, CA: Sage.

van der Mars, H., Rowe, P. J., Schuldheisz, J. M., & Fox, S. (2004). Measuring students' physical activity levels: Validating SOFIT for use with high-school students. *Journal of Teaching in Physical Education*, *23*, 235–251.

Zimmerman, P. H., Bolhuis, J. E., Willemsen, A., Meyer, E. S., & Noldus, L. P. J. J. (2009). The Observer XT: A tool for the integration and synchronization of multimodal signals. *Behavior Research Methods*, *41*, 731–735.