DOCUMENT RESUME

ED 285 917 TM 870 528

AUTHOR Silverman, Stephen; Zotos, Connee
TITLE Validity and Generalizability of Measuring Student Engaged Time in Physical Education.
SPONS AGENCY Texas Univ., Austin. Research Inst.
PUB DATE Apr 87
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; Higher Education; *Physical Activity Level; *Physical Education; Sampling; Test Reliability; *Test Validity; *Time on Task
IDENTIFIERS *Academic Learning Time in Physical Education; Interval Time Sampling; *Time Sampling

ABSTRACT

The validity of interval and time sampling methods of measuring student engaged time was investigated in a study estimating the actual time students spent engaged in relevant motor performance in physical education classes. Two versions of the interval Academic Learning Time in Physical Education (ALT-PE) instrument and an equivalent time sampling method were compared with actual time students spent in motor performance. Subjects were 36 college students enrolled in six different physical education classes: aerobic dance, badminton, basketball, fencing, karate, and volleyball. All instruction was videotaped so that each student could be coded throughout the class. The actual student engaged time and engaged time as estimated by the time sampling instrument were significantly lower than the time as estimated with the two ALT-PE instruments. In addition, a class type by coding instrument interaction occurred indicating that the difference was present for all classes except aerobic dance, where activity is continuous. This result suggests that the popularity of the ALT-PE instruments should be reexamined. (Author/GDC)

VALIDITY AND GENERALIZABILITY OF MEASURING STUDENT ENGAGED TIME IN PHYSICAL EDUCATION[1]

Stephen Silverman
and
Connee Zotos

Department of Physical and Health Education
The University of Texas at Austin
Austin, TX 78712
(512) 471-4406

Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC, April 20, 1987.

[1] This study was supported by a grant to the first author from the University Research Institute, The University of Texas at Austin, Austin, TX.

Abstract

The purpose of this study was investigate whether interval and time sampling methods of measuring student engaged time are valid for estimating the actual time students spend engaged in relevant motor performance in physical education classes. Two versions of the interval Academic Learning Time in Physical Education (ALT-PE) instrument and an equivalent time sampling method were compared with actual time students spent in motor performance. Thirty-six students enrolled in six different physical education classes at a large university were the subjects. All instruction was videotaped so that each student could be coded throughout the class. The actual student engaged time and engaged time as estimated by the time sampling instrument were significantly lower than the time as estimated with the two ALT-PE instruments. In addition, a class type by coding instrument interaction occurred indicating that the difference was present for all classes except where activity is continuous. This result suggests that the popularity of the ALT-PE instruments should be reexamined.

Systematic observation has been used in educational settings to collect data for both research on teaching and to evaluate teaching. The history of systematic observation in classrooms is well documented (Medley & Mitzel, 1963; Rosenshine & Furst, 1973). In recent years much of the focus of systematic observation of teaching has revolved around the use of time (Fisher & Berliner, 1985). As a part of the Beginning Teacher Evaluation Study [BTES, (Fisher, et al., 1978)] Fisher and his associates found that Academic Learning Time (ALT), defined as the amount of time students spend appropriately engaged with the subject matter, was a powerful variable related to student achievement.

As research was being conducted in classroom situations, researchers interested in motor skill were conducting research on teaching in physical education classes. Time as a variable of teacher and student behavior has been one focus of this research in physical education. A widely used instrument to investigate student engagement in physical education is the Academic Learning Time in Physical Education instrument [ALT-PE, (Siedentop, Birdwell, & Metzler, 1979)]. This instrument, since revised and simplified (Siedentop, Tousignant, & Parker, 1982), was developed based on the original instrument for observing classroom processes in elementary school math and language arts classes as part of the BTES study.

Although the original instrument developed for BTES was validated (Fisher, et al., 1978), the two versions of the instrument used in physical education classes have not been validated against actual student engaged time. Validity was assumed based on the system developed with BTES. Physical education, however, occurs in an environment different from the classroom and the new instruments focused on student engagement in movement as opposed to traditional aspects of student engagement in the classroom. The fleeting nature of movement in a physical education class may provide engagement data which are not valid when measured by an interval system such as the ALT-PE instruments. If engaged time as measured by an interval system is not valid for movement settings, policy makers and researchers will then need to interpret data collected with the ALT-PE system with care. It also will be necessary to explore other methods of measuring student processes as an indicator of teacher effectiveness and to collect data for research on teaching. The purpose of this study, therefore, was to compare the two interval recording ALT-PE systems and an equivalent time sampling measuring based on second-by-second recording with actual student engaged time.

Method

Thirty six students enrolled in six videotaped physical education classes were coded for four measures of student engaged time. The two ALT-PE instruments, a time sampling system using the same intervals of time as the ALT-PE instrument, but employing actual second-by-second coding of student time, and actual student engaged time were compared.

Subjects

Students in six elective university physical education classes were the subjects of this study. The six classes represented different subject matters within physical education. The classes were: 1) aerobic dance; 2) badminton; 3) basketball; 4) fencing; 5) karate; and, 6) volleyball. Different physical education subject matter was used to examine student engagement because the typical organization of class practice is different with a different subject matter focus. For example, aerobic dance classes are characterized by continuous movement and karate classes have directions, quick movement, and then feedback.

All students in each class completed a letter of informed consent. During the class period that was observed students wore a numbered pinafore for subsequent identification on videotape. Six students from each class were randomly selected to be coded for the two ALT-PE instruments, the time sampling instrument, and for actual engaged time. A total of 36 students were coded.

Videotape recording

The observed session of each class was videotaped using two cameras and a split-screen generator. One camera had a wide-angle lens and the second camera was equipped with a telephoto lens. This equipment permitted virtually all instruction and practice in each class to be recorded. Second-by-second elasped time was superimposed on the videotape by feeding the video signal through a time-date generator prior to recording on a videocassette recorder (VCR). Teachers wore a cordless, portable microphone and a concurrent audio signal was recorded. The videotape equipment was positioned unobtrusively in a corner of the gymnasium or practice room.

Coding of videotapes for student engagement

Each of the 36 students was coded four times. The first 30 minutes of each class was used for data collection. All engagement coding was performed while viewing the tape on a 23 inch television monitor. The VCR had the ability to replay the tape in a normal, fast, or slow-motion mode.

When the tape was replayed the image displayed on the screen was divided in half horizontally with the image from one camera on the top and the image from the other on the bottom. As students moved from left to right across the top part of the screen they disappeared from the area recorded by the top camera and simultaneously appeared on the left side of the bottom part of the screen.

The two investigators performed all coding for this study. One of the investigators had a great deal of experience using systematic observation for data collection with physical education classes. Prior to collecting data for each coding system, training occurred. The two coders discussed the instrument, coded out loud, coded silently and compared discrepancies, and then coded the same student on separate occasions to determine reliability. Actual data collection did not begin until coders were reliable at a .90 level.

The scored interval method (Johnson & Bolstad, 1973) was used to calculate reliability for each of the ALT-PE instruments. Reliability for the actual engaged time and the time sampling method was calculated using repeated measures ANOVA (Winer, 1971). To ensure that observer drift did not occur during the various data collection phases, additional reliability checks were conducted at the conclusion of each phase. All reliability checks except for one check on the ALT-PE version 1 instrument were above .90. The one reliability check below .90 was .88 indicating that all data collected for the study were highly reliable.

The coding procedures and instrument overview are presented below. The order of presentation is the same order as that which occurred during data collection.

ALT-PE version 1. This interval recording system was developed in the late 1970s (Siedentop, Birdwell, & Metzler, 1979) and patterned after the system developed as a part of BTES. Coding requires the observer to make four decisions during each interval. The first level of decision is the SETTING of instruction as determined by the teacher (e.g. direct instruction, task instruction, etc.). The next level of decision making is CONTENT (whether related to physical education or not). The content related to physical education (e.g. skill practice, scrimmage, knowledge, etc.) is classified. The third level of decision making has been termed LEARNER MOVES and revolves around whether the student is engaged and if the engagement is with motor practice or cognitive knowledge. The final level of decision making is the DIFFICULTY LEVEL (i.e. easy, medium or hard) of the motor or cognitive engagement.

The interval nature of this instrument requires the coder to select one category for each level during observation. For example, a student may not be engaged for most of the interval and then initiate a motor skill at the conclusion of the interval. The hierarchy of coding requires that this interval be coded engaged in motor skill if this occurred during any part of the interval. Therefore, motor engagement is coded for the entire interval even if it occurred only for the last 1/2 second. An assumption of the instrument is that over the course of coding the measurement error will be balanced by longer episodes of motor performance not observed during a coded interval.

During each observation session two of the six students selected from the class could be recorded. The coder observed student one for six seconds and then had six seconds to record the observation. The second student was then observed for six seconds and six seconds were provided to record this observation. The first student then again was observed and the look-record, look-record sequence repeated. In order to maintain the six second cycle the observer listened to an audiotape which paced the coding session by instructing the coder of the interval number and whether observation or recording should be taking place.

ALT-PE version 2. Version 2 of ALT-PE (Siedentop, Tousignant, & Parker, 1982) is a more concise and simplified system of version 1. Coding requires the observer to make two levels of decisions. The first level is the CONTEXT level. Decisions at this level focus on whether the teacher has structured the class for subject matter learning related to knowledge (e.g. knowledge of rules, techniques, etc.), motor performance (e.g. skill practice, scrimmage, etc.) or general non-subject matter activities (e.g. management, transition, etc.). The second level is the LEARNER INVOLVEMENT level. Here the observer decides if the students are or are not engaged (e.g. waiting, off-task, etc.) in motor activity. If the student is engaged in a motor activity the observer records whether the motor activity is appropriate, inappropriate, or performed in a supporting role.

The procedure for coding was the same as for version 1. Two students were observed during the same coding session. These students were coded for the same six second time periods as they were coded during the earlier session.

Actual engaged time. An instrument was developed by the investigators to code the actual second-by-second engaged time of the student. Each student was coded for the entire thirty minutes. Student action was coded into the following categories: 1) motor engagement-appropriate — the student was engaged in motor skill at level of difficulty which was not overly difficult; 2) motor engagement-inappropriate — the student was engaged in motor skill, but the practice was difficult or the student was obviously not successful; 3) supporting — the student was involved in motor activity, but not in active practice of the skill (e.g. feeding a pass to another student practicing the skill, etc.); 4) cognitive — the student was receiving knowledge related to the subject matter from the teacher; 5) waiting — the student was waiting to practice the skill or had completed practice as instructed and was waiting for directions from the teacher; and, 6) management/transition — the student was moving from one area to another, receiving information not related to the subject matter from the teacher, or was retrieving a ball or shuttlecock.

Each of the subjects was coded from the beginning of class until thirty minutes of the class was complete. The observer coded the time a student began in a category, and then the time they moved to another category. This required recording the end of the first category and the beginning of the second category in the appropriate areas on the coding sheet.

After completion of the coding session, the observer subtracted the starting time of each recording from the stopping time and listed the seconds the student spent in that category. Total seconds in each category then was summed. The total seconds spent in the motor engaged-appropriate category was used as the measure of actual student engaged time upon which to compare the other estimates.

Time sampling. In order to determine if six seconds of each 24-second cycle (the observation period of the ALT-PE instruments) was valid for sampling student engaged time, the same six second time periods as used for each student during ALT-PE coding were coded for second-by-second time spent in the categories detailed in the section above. Since the reliability of coding the actual engaged time was very high (above .94 for all reliability checks) and the second-by-second coding for actual time was tedious and time consuming (approximately 100 hours), the data for this category were recorded from the original coding sheet for actual time. The appropriate six second time periods for each student was recorded on a new coding sheet. Therefore, although based on the actual coding for engaged time, the data collected here were for a similar 25% sample of the class as coded earlier with the two ALT-PE instruments. It should be noted again that additional reliability checks were performed on these data and as would be expected all checks were very high (.98).

Data analysis

All intervals of the ALT-PE instruments related to motor engagement at an appropriate level were counted. Since each instrument treats different kinds of behavior differently in the various levels, modifications were made in normal tallying so that all estimates of motor engagement represented the same types of behavior. For ALT-PE version 1, all intervals coded both "motor response" at either an "easy" or "medium" difficulty level were summed. For version 2 , categories coded as "motor engaged appropriate" were summed with those coded "warm-up." The category of "warm-up" was coded as unrelated to subject matter in this version of the instrument, but as a "motor response" in the previous version.

The summed number of intervals of each of the ALT-PE instruments was multiplied by 24 seconds (the total time period represented by one observation). This new value was the estimate of the total number of seconds during the 30 minute (1800 seconds) observation for which the student was engaged in motor appropriate activity. To calculate the estimate for the time sampling instrument the tallied number of seconds was multiplied by four (25% of the 30 minutes was sampled).

The three estimates of engaged time were analyzed with the actual measure of motor appropriate student engagement. A 4 (coding instrument) x 6 (class) ANOVA with repeated measures on the first factor was performed to determine if there were differences among the four measures of engagement and if the type of class interacted with the type of instrument. Significant main effects and interactions were followed-up using the Student-Newman-Keuls post-hoc test and tests of simple effects based on the pooled error term (error mean squared) for the ANOVA.

Results

As noted in Table 1, there were large differences in student engagement as measured by each instrument among the six classes. The same order of mean student engagement by class was evident for each instrument. In addition, smaller standard deviations were found for the classes where students were more likely to perform practice in unison (i.e. aerobic dance and karate).

Insert Table 1 about here

A significant ($p < .001$) main effect occurred for the engagement coding instruments. The mean time for student engagement as determined by the actual second-by-second coding was 667.5 seconds of the 30 minutes (1800 seconds). Of the three methods used to estimate engaged time version 2 of the ALT-PE system provided the highest mean estimate (931.4 seconds). Version 1 of ALT-PE followed very closely with an estimate of 931.3 seconds. Time sampling based on the second-by-second coding provided an estimate of student engagement nearly identical (658.7 seconds) to the actual mean engagement time. The follow-up test indicated that the estimates of the two versions of ALT-PE were significantly higher than the actual time and time sampling measurements.

A significant ($p < .001$) class by engagement category interaction was present. For five of the classes (all but aerobic dance) both ALT-PE estimates were higher than time sampling or the actual estimate. For students in the aerobic dance classes, the four measures of engaged time were statistically equivalent.

Discussion

It is clear that despite the frequent use of the ALT-PE instruments in research on teaching in physical education (Dodds & Rife, 1983) that they do not provide valid measures of actual engaged time. As Shavelson, Webb, & Burstein (1986) have pointed out, often instruments that appear to have face validity will, when used, record behavior which is inconsistent with actual behavior. In this instance the assumption that measurement error created by assigning an entire interval to motor engaged time will be "evened out" by more continuous movement in other intervals is not true except for the aerobic dance classes.

The interaction of the class with the coding instrument showed that for all classes but aerobic dance the ALT-PE estimates of engaged time were higher than the time sampling estimate or actual time. This makes sense when considering the way in which instruction is conducted in the different classes. In aerobic dance virtually all intervals are completely filled with engaged time. Movement is continuous throughout. In the other activities, however, movement was often of short duration such as a pass in volleyball, a clear in badminton, or two cycles of a kick in karate. In the five classes where the differences existed among instruments, there were very few instances where movement was of long-term duration.

The time sampling instrument provided a valid estimate of student engagement. In fact, the sampling of 25% of the lesson in systematic cycles of six seconds found an engaged time estimate nearly identical to the actual time for each class and for the combined data set. It seems reasonable that systematic sampling in this manner, where a second-by-second recording is occurring, would provide such data. As has been noted before by many others (for instance, Kerlinger, 1973) that observations which are not random, and are less systematic (arbitrary or convenient selections of time) may not provide equally valid data even if actual time spent in certain categories is coded. This may be particularly true in physical education classes where it is likely that practice takes place in spurts; warm-up, demonstration, practice feedback, etc. It would be helpful to investigate if other, less tedious time sampling methods would provide equally valid data.

Although the two ALT-PE instruments may not be valid measures of actual engaged time, the rank order of engaged time as measured by all four instruments is identical. If policy makers, evaluators, or researchers are interested in only if engaged time is more or less than some other time or situation, as measured by the same instrument, it appears that this may be proper. When used in this way it will be necessary to refrain from invalid comparisons with other instruments.

The use of systematic observation of teaching will continue for obtaining data for research on teaching (Brophy & Good, 1986; Shulman, 1986) and to evaluate teaching performance (for example, Capie, Anderson, Johnson, & Ellett, 1979; Texas Education Agency, 1986). As has been shown through this study, instruments which are appropriate in one setting may not be validity generalized to other settings. If research, evaluation, and policy decisions are to be valid, the selection of the observation instrument must be done with care.

References

Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), Handbook of research on teaching [(3rd ed.) pp. 328-375]. New York: Macmillian.

Capie, W., Anderson, S.J., Johnson, C.E., Ellett, C.D. (1979). Teacher performance assessments instruments: A handbook for interpretation. Athens, GA: Teacher Assessment Project, University of Georgia.

Dodds, P., & Rife, F. (Eds.). (1983). Time to learn in physical education: History, completed research, and potential future for academic learning time in physical education. Journal of Teaching in Physical Education, Monograph 1.

Fisher, C.W., & Berliner, D.C. (Eds.). (1985). Perspectives on instructional time. New York: Longman.

Fisher, C.W., Filby, N.N., Marliave, R., Cahen, L.S., Dishaw, M.M., Moore, J.E., & Berliner, D.C. (1978). Teaching behaviors, academic learning time and student achievement: Final report of Phase III B, Beginning Teacher Evaluation Study (technical report V-1). Washington, DC: The National Institute of Education.

Johnson, S., & Boltad, O. (1973). Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. Hamerlynk, L. Handy, & E. Mash (Eds.), Behavior change: Methodology, concepts, and practice. Champaign, IL: Research Press.

Kerlinger, F.N. (1973). Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Medley, D.M., & Mitzel, H.E. (1963). Measuring classroom behavior by systematic observation. In N.L. Gage (Ed.), Handbook of research on teaching (pp. 247-328). Chicago: Rand McNally.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In M.W. Travers (Ed.), Second handbook of research on teaching (pp. 122-183). Chicago: Rand-McNally.

Shavelson, R.J., Webb, N.M., & Burstein, L. (1986). Measurement of teaching. In M.C. Wittrock (Ed.), Handbook of research on teaching [(3rd ed.) pp. 50-91]. New York: Macmillian.

Shulman, L.S. (1986). Paradigms and research programs in the study of teaching. In M.C. Wittrock (Ed.), Handbook of research on teaching [(3rd ed.) pp. 3-36]. New York: Macmillian.

Siedentop, D., Birdwell, D., & Metzler, M. (1979, April). A process approach to measuring teacher effectiveness in physical education. Paper presented at the Annual Conference of the American Alliance of Health, Physical Education, Recreation, and Dance, New Orleans.

Siedentop, D., Tousignant, M., & Parker, M. (1982). Academic learning time - physical education, 1982 revision: Coding manual. Columbus, OH: School of Health, Physical Education and Recreation, The Ohio State University.

Texas Education Agency. (1986). Teacher appraisal system: Teacher oreientation manual. Austin, TX: Author.

Winer, B.J. (1971). Statistical principles in experimental design (2nd. ed.). New York: McGraw-Hill.

Table 1

Instrument by Class Means and Standard Deviations for
Seconds of Student Motor Appropriate Engagement

	ALT-PE version 1	ALT-PE version 2	time sampling	actual time
Aerobic Dance	1639 (55)	1604 (10)	1614 (10)	1615 (3)
Badminton	932 (167)	883 (248)	501 (121)	513 (131)
Basketball	334 (129)	318 (73)	244 (51)	220 (59)
Fencing	749 (77)	756 (111)	623 (77)	619 (74)
Karate	1329 (76)	1471 (50)	701 (20)	766 (8)
Volleyball	604 (93)	592 (116)	268 (118)	271 (100)
Total	931 (456)	937 (480)	659 (472)	668 (476)

NOTE: Values are rounded to the nearest whole second.