DOCUMENT RESUME

ED 198 173 TM 810 141

AUTHOR Karweit, Nancy: Slavin, Robert E.
TITLE Time-On-Task: Issues of Timing, Sampling and Definition.
PUB DATE Oct 80
NOTE 22p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement: *Classroom Observation Techniques: Correlation: Elementary Education: Research Design: *Sampling: *Time Factors (Learning)
IDENTIFIERS Comprehensive Tests of Basic Skills: *Time On Task

ABSTRACT

This paper addresses four issues in the design and execution of behavioral observation in classrooms. These four issues relate to the consequences of using different observation intervals, schedules of observation, student sampling methods, and definitions of on-task and off-task behavior for reliability, means, and correlations of time on-task and achievement. A field study observed 108 students in 18 elementary classrooms. Pre and post-achievement data were also collected. The data permit simulations of different intervals, schedules, sampling methods, and definitions for determination of their effects on the outcomes of behavioral observation. Findings suggested that: (1) altering definitions of time-on-task to include momentary off-task behaviors affected the conclusions for the importance of time-on-task: (2) sampling segments of instruction would tend to obscure the positive results for time-on-task: (3) reducing the number of days of observation also weakened the effects of time-on-task: (4) timing of the observation was not very important for the noted effects: and (5) reducing the number of students to less than six may adversely affect reliability. (Author/RL)

Time-On-Task: Issues of Timing, Sampling and Definition

Nancy Karweit
Robert E. Slavin

Center for Social Organization of Schools
The Johns Hopkins University
Baltimore, MD 21218

October, 1980

Time-On-Task: Issues of

Timing, Sampling and Definition

October 1980

Abstract

This paper addresses four issues in the design and execution of behavioral observation in classrooms. These four issues relate to the consequences of using different observation intervals, schedules of observation, student sampling methods, and definitions of on-task and off-task behavior for reliability, means, and correlations of time on-task and achievement. A field study observed 108 students in 18 elementary classrooms. Pre and post achievement data were also collected. The data permit simulations of different intervals, schedules, sampling methods, and definitions for determination of their effects on the outcomes of behavioral observation.

Introduction

Research interest has recently focused on the centrality of time-on-task for understanding classroom effects and effectiveness (Fisher et al., 1976a, 1976b; Filby and Marliave, 1977; McDonald and Elias, 1976; Cooley and Leinhardt, 1978). This research has provided important evidence that links classroom practices, time-on-task and learning outcomes. Although the evidence in general points to positive and meaningful effects of time-on-task, the results are not consistent across studies nor across grade levels/ subject matters within studies (e.g., the results obtained in the Beginning Teacher Evaluation Studies, BTES, for mathematics/reading at grades 2/5). Moreover, the effects documented for time-on-task, although positive, have not been uniformly large.[1] Nonetheless, the effects for time factors have assumed appreciable stature by virtue of the fact that time factors can be altered, whereas more statistically important factors, such as family background or entering aptitude, are difficult or even impossible to alter.

Thus, the use of time in classrooms continues to be a central theme in educational research. The fact that the results are modest and inconsistent has been attributed to particular methodological or research design problems, not problems with the assumptions guiding the research. That is, the assumption that classroom practices have appreciable impacts on time-on-task which in turn affect the degree of learning is generally not at issue. The present state of encouraging but not entirely clear results is taken to indicate the existence of methodological as opposed to theoretical problems.[2]

Given this slant on the problem, it seems reasonable to ask to what extent the nature of the present findings are due to particular methodological choices or decisions. In particular, it seems useful to explore how the observation scheme used, the timing of the observation, the length of the observation and the number of observations may affect the detection of time-on-task effects. This paper addresses these issues by using an existing set of observational data and manipulating it to conform to alternative sampling, timing and definitional choices. We examine the alternate effects of choices in 5 areas:

1. definition of off-task behavior.
2. length of observation visit
3. days of observation
4. scheduling of observation
5. sampling of students for observation

Data

The data were collected in four elementary schools in a rural Maryland school district. Subjects were students in grades 2-5 in 18 classes taught by 12 teachers. All students were pre- and post-tested in February 1978 in reading, language arts, math and social studies using the Comprehensive Test of Basic Skills. Students in each class were assigned to the top third, middle third, or lower third of the class based on the pre-test information, and two students (one boy and one girl) were chosen from each third for observation. The observations were thus conducted for six students per class, 108 total students through the second semester of 1978, and the post-test was given in May, 1978.

Students were observed during their mathematics classes, which averaged 50 minutes. Each classroom was observed for at least nine days, and some for as many as twenty-one days. The observers recorded three pieces of information for all six students during a thirty second interval; the nature of the task (procedural, seatwork, or lecture); the student's response to the task (on-task, off-task or no task opportunity), and the content of the instruction (e.g., two digit multiplication, or going over p. 147).

All six students were observed in a predetermined order every thirty seconds. To determine on- or off-task behavior, the observer took a quick look at the student's behavior and recorded the response at that particular instant. The observers were trained not to dwell on deciding whether a behavior were on- or off-task, but to record their first impression in accordance with established definitions of on- and off-task responses.

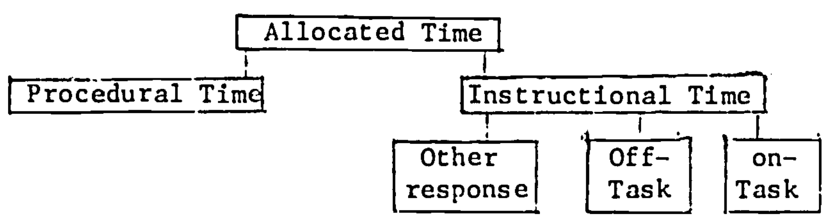
On average, 100 observations per day were recorded for each student, detailing the task, the content of instruction, and the response. Counting all the daily observations, we logged on average about 1000 observation points for each student in the sample, or about 110,000 observation points total.

Because of the size of the data base, we entered the task, content and response codes in a summary form which maintained the essentials of the information. Each entry pertained to a specific task or activity and gave the number of seconds each child was on-or off-task during that time. For example, if the class were involved in seatwork during the first ten minutes of the class and then the teacher explained the seatwork during the next eight minutes, we created two entries, one detailing the on/off task behavior during seatwork, and the other giving the on/off task behavior for each of the six children during the teacher-directed activity. From these data, a "day" record was constructed which summarized the daily task, content and response patterns for each child.

In addition, a special data set containing each 30 second record of task, content and response was compiled for five of the eighteen classrooms. These supplemental data will be used along with the basic data in the analyses.

Definition of On- and Off-Task

On- and off-task behaviors were coded during instructional portions of the lesson only. However, a child could also have a response other than on- or off-task during instructional time. The diagram below depicts the different categories and when they could occur in this observational scheme,



The allocated time was the clock time scheduled for the mathematics class. Procedural time included any time spent lining up, receiving instructions, being involved in disciplinary action, going to fire drills, being interrupted by the P.A. system and the like. Instructional time pertained to the time spent specifically on mathematics instruction; discussion of world events, elections, snow storms and other material not pertaining to math was not coded as instructional time. On-task behavior was defined as behavior appropriate to the task at hand. The definition of appropriate behavior depended upon the task and specific rules of the classroom. "Other response" was used to cover situations in which the child was not on-task but was not off-task either. Such situations arose when the child was sharpening a pencil, walking to another part of the room to obtain new materials, waiting for the teacher to help with a problem, or doing some other activity because the original assignment was finished.

We focused on two particular problems in assessing off-task behavior. One involved the effect of including momentary off-task behavior; the other involved the effect of including no-task-opportunity time (i.e., "other response") as off-task behavior.

a. Momentary off-task behavior

During any class period, children may gaze out the window, fidget, or otherwise be momentarily distracted. On the one hand, this momentary off-task time can be looked upon as insignificant for the learning process. On the other hand, momentary off-task behavior may be signalling declining attention and motivation and might therefore be important for understanding the learning process. The issue is whether these flickers of inattention should be treated as measurement noise and thus ignored or as true indications of the underlying variable of interest, engagement with learning. The final decision in this matter depends upon conceptualizations of "engagement"; here, we simply examine the measurement consequences of the choice made. We simulated the "noise" decision by changing all off-task behavior of less than one minute to on-task behavior in the supplemental sample of five classrooms. The average rate of on-task behavior increased from .79 to .83 while the standard deviation was reduced from .08 to .07. Including the momentary off-task behavior yielded correlations of .24 between on-task and pre-test score and .45 with post-test score. Excluding momentary off-task behavior, these correlations became .33 and .29. We carried out regressions of post-test on pre-test and the alternative measures of on-task behavior. Using the measure which excluded momentary off-task behaviors produced more modest results ($p < .10$) than did using the more inclusive measure ($p < .05$).

Whether to include momentary inattention or not should be based on the particular model of learning one has formulated. Certain views of the learning process may be compatible with inclusion of these momentary distractions; other views would not be. The present exercise was not intended to shed light on whether a particular point of view is proper or improper, but to illustrate that the methodological decision to include/exclude these flickers of inattention affected the results obtained.

b. Other response and off-task behavior

The dichotomy of on- or off-task provides a working categorization of student responses to instruction, but there are numerous ambiguous situations in which the student is not on-task, yet could not be considered off-task. For example, a student may have finished an assignment and have nothing more to do. Students who finish early are likely to be those who need less time, i.e., have more aptitude for the particular task at hand; thus the amount of finished time should be positively related to achievement, in contrast to the negative relationship of off-task variables and achievement. In our data, the correlation between finished time and post-test score was .19 while the correlation between off-task and post-test score was -.28. In regressions (not detailed here) in which finished time was included with off-task time, the effects of off-task time were diminished appreciably.

Length of observation visit

An important design consideration is the length of the observation period. One could observe a single classroom all day long, for some fixed fraction of the day, or for some specific instructional program. Or, a combination of these lengths of observation might be used. Because our interest was in how the use of time affects mathematics achievement, we observed students during their entire mathematics instruction. It was not possible (given our budget constraints on ovserver time) to observe all teachers within a school. An alternate decision would have been to observe more teachers, but for some smaller segment of their mathematics instruction. We might have decided, for example, that instead of visiting one teacher for sixty minutes we might have used one of the combinations below:

NO. TEACHERS	NO. MINUTES	TOTAL TIME
2	30	60
3	20	60
6	10	60

The choice among these alternatives is basically between getting enough classrooms to provide stable estimates of the effect of time-on-task, and scheduling sufficient time to ensure that the observed behavior is represenatative. If time-on-task is distributed fairly uniformly across the day or the period of instruction, then a time sample may be entirely adequate. Table 1 gives the means and standard deviations of time-on-task for nine days of observation in one classroom. The first columns provide statistics for the first 10 minutes of class; the

Table 1 About Here

second for the first 20 minutes, and the third for the first 30 minutes. The overall mean for the time period is supplied as well as the mean for the particular 10 minute segment. The average on-task time in this class was markedly higher during the first 10 minutes of instruction than it was for the next 10 or 20 minutes. Clearly, the timing of observation in this classroom was important for the results obtained as time-on-task was not distributed evenly across the mathematics class time. Other classes started off with lower on-task rates, seemed to warm up to instruction, and have higher on-task rates, and then to die down. Still other classrooms had no consistent pattern at all. Consequently, it is difficult to predict what the effect in general would be if selected portions only of the class time were observed. Thus, although the the effect of observing shorter periods may not be consequential for the reliabilities obtained (see Rowley, 1976), how those periods are selected may be very consequential.

To illustrate this point, we regressed post-test achievement scores on pre-test scores and alternate measures of on-task rate, namely measures from the first ten, twenty, thirty and fifty minutes of instruction. The F values obtained for the time-on-task measures were .010, 1.22, 3.09 and 4.34, respectively. The n of this sample was extremely small (22 students): however, the results suggest that

Table 1

On-Task Rate for Selected Portions
of Mathematics Instruction
in one classroom

Day	Minutes 1 - 10		Minutes 1 - 20		Minutes 1 - 30	
	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ
1	.906	.066	.878	.046	.865	.051
2	.911	.086	.815	.059	.805	.092
3	.922	.078	.923	.079	.921	.084
4	.739	.236	.818	.062	.775	.062
5	.817	.002	.690	.158	.653	.195
6	.889	.087	.869	.073	.804	.099
7	.884	.169	.866	.188	.809	.175
8	.958	.066	.928	.063	.889	.088
9	.825	.196	.841	.148	.850	.171
$\overline{X}$	.872		.848		.819	
$\overline{X}$ (this segment)	.872		.824		.761	

observing for shorter segments would have appreciably altered the effects obtained.[3]

Altering the number of days of observation

Conventional wisdom has it that about ten days of observational data should be sufficient to accurately portray the activities of a classroom. However, few studies have investigated the effects of observing classrooms for fewer or more days, even though this question is of considerable design and practical importance. If we can obtain sufficient information in a shorter period (e.g. five days instead of ten), it would be possible to observe substantially more classrooms without appreciably altering the observation costs.

In the present data set, we observed some classrooms for as many as 21 school days and others for as few as 9 days. With these data, then, we can pretend that we had observed a fixed number of days (e.g. 3, 4, 5, 6, 7, 8, 9) and assess how this observation schedule would have affected the detection of effects of time-on-task on achievement. We think of time-on-task as a variable which is influenced not only by an individual child's disposition, aptitude, and idiosyncracies, but also by the instructional setting in the class and by external events such as the daily weather. Each child may have a stable rate of on-task behavior with daily fluctuations depending upon his response to the classroom and other environmental settings. Given this view of time-on-task as a variable, a natural way to capture the daily and individual variation is to view each day's time-on-task as an item in a scale of total time-on-task. We can then see how consistent the behavior is across a differing number of days or items in the scale.

As expected, increasing the number of days does provide an overall increase in reliability. The median coefficient alphas obtained for 3-9 days were .54, .57, .71, .73, .79, .81, .82. Whether the increase in reliability obtained from observing nine days vs. 5 days is consequential depends on the effect one is trying to document. Because reliability determines the maximal correlation that one can find between achievement outcomes and time-on-task, the obtained reliability is of some consequence. To assess the effect of these variations in reliability, we used the third grade sample (n=36) and regressed post-test CTBS score on pre-test and alternate measures of time-on-task, namely measures obtained from:

1. five days observation
2. nine days observation
3. eighteen days observation

Table 2 shows that had we observed for the first five or first nine days our effects for time-on-task would have been much more modest. The

Table 2 About Here

"18-day" results show significant effects for on-task minutes, engagememt rate and off-task rate. Had we observed the same classrooms, but for fewer days, the results obtained would have been much weaker:

It is possible that the days at the end of the observation were significantly different from the days at the beginning; if this were the case we would be witnessing an effect for timing and not for length. This issue is explored in the next section.

Table 2

Comparison of Results Obtained
for time-on-task using 5, 9, and 18
days of observation

	18 days		9 days		5 days	
	b/beta	F	b/beta	F	b/beta	F
time-on-task rate	47.51 (.178)	6.56 $p < .05$	33.62 (.129)	3.21 $p < .10$	32.25 (.138)	3.62 $p < .10$
time-on-task rate	.249 (.165)	5.14 $p < .05$	.156 (.111)	2.28 n.s.	.131 (.11)	2.19 n.s.
time-off-task rate	-48.1 -(.147)	4.33 $p < .05$	-32.9 -(.10)	2.07 n.s.	-.109 -(.03)	.141 n.s.
time-off-task minutes	-.450 -(.121)	2.86 n.s.	-.329 -(.09)	1.39 n.s.	-16.76 -(.05)	.459 n.s.

Timing of observation days

Throughout the school year, there are no doubt more intensive and less intensive periods of instruction. At the beginning of the school year, for example, much of the instructional time may be spent in review or in establishing classroom rules and norms for behavior. In many urban schools, with high rates of student mobility, the first six weeks are needed to stabilize the school enrollment. Because of the constant transferring in and out of classrooms, this early part of the semester is often an instructional loss. Other examples of uneven distributions of effort throughout the school year are the days immediately before and after major holidays, such as Christmas and Easter vacations. These sources of differences in time-on-task are predictable and probably similar from class to class. In addition, there may be different levels of seriousness in the classroom, depending upon the amount of material the teacher has covered and the amount she expected to cover by that point in time. This variation in the teacher's expectation for levels of attentiveness would then not likely be the same from class to class.

We are able, in a limited fashion, to see if time on-task differs by time of year, using these data. For four classrooms we observed students for a ten day period in February and also in May. The means and standard deviations for these classrooms are provided in Table 3 for the two different time periods. Table 3 also provides the reliabilities for the two periods of observation (column 5 and 6) and for two mixed scales S1 and S2 composed of equal number of items from February and May. The reliabilities and the means do not appear to be very different for the two time points. This table supplies limited evidence of the consistency of the classroom over time, which suggests that the timing of the observational period may not be all that consequential. It also suggests that our failure to find significant

Table 3

Comparison of Mean Values and reliabilities obtained or time on task in February and May

Feb. Means	May Means	Feb. α	May α	Combined α	S1 α	S2 α
.844	.856	.92	.96	.97	.93	.94
.899	.900	.76	.72	.71	.70	.53
.929	.930	.67	.76	.85	.70	.70
.842	.847	.85	.76	.79	.56	.71

effects for time-on-task using only nine days of observational data was most likely due to the decreased reliability of the scale and not due to scheduling effects.

Altering the number of students sampled in the classroom

Another decision which has to be made is whether to observe all students in the room or to follow a sample of students. Whether to observe the entire classroom or selected students depends largely on the purpose of the observation. If one is interested primarily in how classroom organization affects time-on-task, the entire class would probably be observed. Other strictly pragmatic elements such as high absenteeism or sensitivity of identifying students for observation may influence this decision.

Given that the practical and theoretical concerns the dictate that sampling should take place, the question is how many students are needed to obtain a reliable estimate of the on-task behavior for the class. We can examine this issue in two ways with these data. In one classroom, we actually observed twelve students as opposed to six, and comparing the class means and standard deviations and reliability obtained for these six vs. twelve shows them to be very similar ($x_{12} = .87$, $x_6 = .86$, $r_{12} = .92$, $r_6 = .89$). Another way we can focus on this issue is by reducing the number of students and comparing the obtained reliabilities. We used a random selection of three of the six students to assess the effect this sampling might have on reliability. The median reliabilities were not appreciably reduced by selecting only three students.[4] However, given the fragility of time-on-task effects which we have documented here, it would seem worthwhile to keep reliability as high as possible. In this instance, observing six students would seem desirable.

Summary and discussion

This paper has examined how various methodological decisions may influence studies of the effect of time-on-task on achievement. We found that altering definitions of time-on-task to include momentary off-task behaviors affected the conclusions for the importance of time-on-task. We found clear evidence that sampling segments of instruction would tend to obscure the positive results for time-on-task. We further showed that reducing the number of days of observation also weakened the effects of time-on-task. The timing of the observation was not very important for the noted effects, however. Finally, we explored the effect of sampling differing numbers of students, and suggested that reducing the number of students to less than six may adversely affect reliability.

The findings in this paper suggest that although there is an understandable urge to lessen the observation time in order to bolster the number of settings observed, such steps should only be taken cautiously. Whether the effects detected and not detected here are bound up with the particulars of this observation study can only be determined by more systematic examination of these methodological issues. In this sense, we hope the paper serves more as a source of what the question might be than of what the answer is. What this paper does show is that methodological decisions, including some that appear quite minor, can have major consequences for the conslusions that are drawn from observational data.

Notes

1. A typical finding has been that time-on-task when added to a regression of post-test on pre-test will increment R^2 by about 3 percent. Although increments to R^2 provide a conservative view of the importance of a variable, other indicators, such as the magnitude of the beta weight or the residual variance accounted for have not been substantial either.

2. An alternative perspective would be that the work is basically atheoretical so that it is natural to fault the methodology.

3. For five of the eighteen classrooms, we coded each 30 second interval of task, content and response. From this sample, the twenty-two students who had complete test and observational data were used in the regressions reported in this section.

4. The median reliabilities obtained for three students in comparison to six students for three to nine days of observation are:

	3 days	4 days	5 days	6 days	7 days	8 days	9 days
3 Students	.43	.65	.63	.63	.71	.77	.81
6 Students	.54	.57	.71	.73	.79	.81	.82

References

Cooley, W. C. and Leinhardt, G. The Instructional Dimensions Study: The Search for Effective Classroom Processes. Learning Research and Development Corporation, Pittsburgh, 1978.

Filby, N. N. and Marliave, R. Descriptions of Distributions of ACT within and across Classes During the A-B Period. (Technical Report IV-1a) Far West Laboratory for Educational Research and Development, 1977.

Fisher, C. W., Filby, N. N., Marliave, R. S., Cohen, L. S., Moore, J.E. and Berliner, D. C. A Study of Instructional Time in Grade 2 Mathematics (Technical Report II-3). San Francisco, Calif.: Far West Laboratory for Educational Research and Development, 1976a.

Fisher, C. W., Filby, N. N., Marliave, R. S., Cohen. L. S., Moore, J. E. and Berliner, D. C. A Study of Instructional Time in Grade 2 Reading (Technical Report II-4). San Francisco, Calif.: Far West Laboratory for Educational Research and Development, 1976b.

McDonald And Elias, Beginning Teacher Evaluation Study, Phase II 1973-74. Princeton, NJ: Educational Testing Service.

Rowley, G. L. The Reliability of Observational Measures. American Educational Research Journal, 1976, 13, 51-59.