DOCUMENT RESUME

ED 257 830 SP 026 248

AUTHOR Anderson, Lorin W.; Mandeville, Garrett K.
TITLE Classroom Activity Structures and the Generalizability of Teacher Behavior.
PUB DATE [85]
NOTE 21p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Techniques; Elementary Education; Evaluation Methods; Generalization; *Instructional Design; Learning Strategies; *Teacher Behavior; *Teacher Evaluation; Teaching Methods

ABSTRACT

The purpose of this study was to examine the influence of several instructional formats (e.g., lecture, discourse, seatwork) on the generalizability of teacher behaviors. Two structured observation instruments were used to observe two samples of teachers: 42 fifth grade science teachers on eight occasions, and 87 fifth grade mathematics teachers on six occasions. The first instrument provided information pertaining to the instructional format; the second yielded data on specific teacher behaviors. As hypothesized, the generalizability of teacher behaviors within instructional formats was greater than that across formats. However, the influence of instructional formats on the generalizability of teacher behaviors was greater in science than in mathematics. (Author)

Classroom Activity Structures and the Generalizability of Teacher Behavior

Lorin W. Anderson

Garrett K. Mandeville

University of South Carolina

Abstract

The purpose was to examine the influence of several instructional formats (e.g., lecture, discourse, seatwork) on the generalizability of teacher behaviors. Two structured observation instruments were used to observe two samples of teachers; 42 fifth grade science teachers on eight occasions, and 87 fifth grade mathematics teachers on six occasions. The first instrument provided information pertaining to the instructional format; the second yielded data on specific teacher behaviors. As hypothesized the generalizability of teacher behaviors <u>within</u> instructional formats was greater than that <u>across</u> formats. However, the influence of instructional formats on the generalizability of teacher behaviors was greater in science than in mathematics.

During the 1970s the focus of the American public in the field of education was clearly on the student. Basic skills tests, functional literacy tests, minimum competency tests, and proficiency tests were mandated by legislation passed in the vast majority of the fifty states. If the 1970s can be referred to as the decade of student assessment and evaluation, the present decade is rapidly becoming the decade of teacher assessment and evaluation. An increasing number of states, including almost all of the Southeastern states, have legislated mandatory classroom observations for the purpose of evaluating teachers.

In many of these states these observations provide the primary source of data on which decisions concerning a teacher's employment, reappointment, and promotion are based. Unfortunately, the required number of observations per teacher is quite small. Kowalski (1978) surveyed administrators in 375 school systems concerning their current policies and procedures governing teacher evaluation. She found that the maximum number of times any teacher was required to be observed in a single year was 3; the typical number being once a year (or less) for tenured teachers and twice annually for untenured teachers.

The increased reliance on observations of teacher performance in teacher evaluation raises the obvious question: to what extent are the data collected from two or three observations sufficiently reliable for sound and defensible evaluations of teachers? The available evidence suggests that such data are not sufficiently reliable. Erlich and Shavelson (1978), for example, found that more than 10 observation occasions would be needed to achieve a generalizability coefficient of 0.70 for approximately three-fourths of the teacher behaviors included on a popular observation instrument. Quite clearly, ten observations of individual teachers in a local school district is impractical.

On the surface, then, local school administrators seem to be facing a major quandry. On the one hand, they are required to base their evaluations of teachers largely on classroom observations. On the other hand, because of practical constraints the number of observations that can be made of individual teachers does not yield sufficiently reliable data to do so.

Fortunately, however, the results of recent studies of teachers suggest that the lack of generalizability is not the result of random variation in teachers' use of particular behaviors, nor the misperceptions or misrecording of observers. Rather, the lack of generalizability of teacher behaviors across occasions seems related to a variety of so-called "contextual" factors or variables.

Evertson and Veldman (1981), for example, found that certain teacher behaviors are more or less likely to be exhibited at particular times of the year. Stayrook and Crawford (1978) supported this finding in an experimental study. In fact, Stayrook and Crawford found that time of year accounted for more variation in teacher behavior than did the treatment.

Evertson and Veldman also found that different teachers teaching different subject matters (i.e., mathematics and English) exhibited quite different behaviors in their classrooms. Stodolsky (1984) extended this result when she found that the same teachers teaching different subject matters (i.e., fifth grade mathematics and social studies) exhibited very different behaviors. In combination these studies suggest that contextual variables such as the time of year during which the observations are being conducted, and the subject matter being taught influence the generalizability of teacher behaviors.

Both time of year and subject matter, although interesting and potentially useful for enhancing our understanding of the limits of generalizability of teacher behavior, are distal context variables. That is, they are variables outside of the classroom and, hence, outside of the control of the teachers. What appears to be needed in an effort to further our understanding of the limits of teacher behavior are more proximal context variables; that is, variables that are "inside" the classroom and under the control of the teachers. Stodolsky (1984) suggests that the variable "activity structure" may be one of the more promising of these proximal context variables. The general purpose of this study was to examine the influence of one dimension of activity structure, namely, instructional format, on the generalizability of teacher behavior.

Activity Structures

The concept of activity structure is derived from the field of ecological psychology. Pioneers in this field included Roger Barker, Paul Gump, and Jacob Kounin. As defined by Stodolsky (1984) the activity structure of a classroom consists of "the various activities taking place. ... (it) includes the salient aspects of the physical environment, a cataloguing of the persons who are present, and the main tasks or types of activities in which the children and teacher are participating" (p. 14). Three of the most important features or dimensions of an activity structure are instructional format (that is, the general arrangment in which instruction is delivered), pacing (that is, who is in charge of "moving things along"), and cognitive level (that is, the type of intellectual processes embedded in the goal or objective of the activity). Of these three features or dimensions the one most immediately recognizable by an observer in a classroom is the instructional format. As a consequence, the instructional format was selected to be used as the focal point for our investigation of the generalizability of teacher behavior.

Instruments

The data to be reported were collected from two instruments, both modifications of instruments developed by Stallings (1977). The first instrument, the Snapshot, was used to record the instructional format. Although eight instructional formats were included on the Snapshot, several were eliminated from the analysis because of their low frequency of occurrence. In the sample of science teachers, for example, only 17 of the 42 teachers assigned written seatwork. Furthermore, the median number of written seatwork segments for these 17 teachers was 1. As a consequence, written seatwork was eliminated from the analysis for the sample of science teachers.

In the sample of mathematics teachers the instructional formats of discourse and laboratory seatwork were eliminated for the same reason. Finally, in both samples the instructional formats of review, testing, reading seatwork, and oral practice almost never occurred and were eliminated. As a result of these eliminations two formats (lecture and written seatwork) remained for the mathematics sample and three formats (lecture, discourse, and laboratory seatwork) remained for the science sample.

Two forms of the Snapshot were used by the observers. With the first form the observers coded the instructional format pertaining to each group of students in the classroom. [In the event of whole-class instruction only one instructional format was coded]. With the second form the observers coded the instructional format pertaining to eight, randomly preselected students. [Again, in the event of whole-class instruction only one instructional format was coded for all students]. Since virtually all of the instruction observed in the classrooms included in this study was of the whole-class variety, the above distinctions are somewhat academic.

The second instrument, the Five-Minute Interaction (FMI), was used to record the teacher's display of specific behaviors. The behaviors were arranged into several categories. Six categories of behaviors will be used to report the results of the study: 1) instructional, 2) questioning, 3) responses to questions, 4) feedback, 5) classroom management, and 6) silence, or the absence of verbal interactions. Observers coded the nature of the observed teacher behaviors and teacher-student interactions every five-seconds or each time the behavior or interaction changed. Thus, approximately 60 codes were made in the five minutes during which the FMI was used (that is, 12 codes per minute times 5 minutes).

Sample

The results pertaining to two samples of teachers will be presented. The first sample consists of 42 fifth-grade science teachers. Teachers in this sample were observed on eight occasions during the school year. The second sample consists of 87 fifth-grade mathematics teachers. Teachers in this sample were observed on six occasions during the school year. Teachers in both samples taught in countries located in Southeast Asia.

Procedures

During each observation, the instruments were employed in a fixed sequence: first, the Snapshot; then, the FMI; and finally, a modification of the Snapshot where the focus of the observer was on individual students rather than the whole group. Each sequence required approximately eight minutes, and each observation period was approximately 40 minutes in length. As a consequence, five sequences were obtained during each observation period. Furthermore, the data were initially aggregated to the eight-minute level.

Each sequence was examined separately. First, the instructional format codes on the Snapshot and the modified Snapshot were considered. If the instructional format codes were identical, the assumption was made that the teacher behaviors as coded on the FMI occurred within a single instructional format. If the instructional format codes were not identical, the FMI data within that sequence were excluded from further analysis.

Second, frequencies of teacher behaviors within each eight-minute segment were computed for the FMI data that were retained. If the instructional formats of two or more consecutive segments were identical, the frequencies of the teacher behaviors in these segments were added since it was assumed that all of the teacher behaviors coded in these segments occurred within a single, continuous instructional format.

Finally, these frequencies were converted to percents by dividing the frequency of occurrence of each behavior by the total number of behaviors coded during that segment. Twenty behaviors and interactions associated with the six aforementioned categories were retained for further analysis. These behaviors were those most closely resembling ones typically included on observation instruments used for the purpose of teacher evaluation.

Within each instructional format, intraclass correlations were computed for each of the 20 behaviors and interactions. Intraclass correlations also were computed for all behaviors independent of the instructional format within which they were exhibited. Based on these intraclass correlations the number of observations needed to achieve an intraclass correlation coefficient of at least 0.70 for each behavior or interaction was estimated. A minimum coefficient of 0.70 was selected because this value had been used in related prior research (Erlich and Shavelson, 1978) and it seemed reasonable to consider such a value minimally acceptable for decision-making purposes.

Results

Tables 1 and 2 display the mean percents of behaviors and interactions displayed within each instructional format and across all formats for the fifth grade science and fifth grade mathematics teachers, respectively. As can be seen in these tables a number of the behaviors and interactions were used very infrequently by teachers. Five of the 20 behaviors and interactions had frequencies of occurrence less than 1 percent of the total number of interactions in both samples (uses examples, asks opinion questions, asks "do you understand," says "wrong,"

Table 1
Mean Percentages of Behaviors Recorded
(Fifth Grade Science; n = 42; obs=8)

Behavior	Total	Lecture	Discourse	SW/L
Teaching				
Explains	7.59	8.59	9.98	4.30
Explains with Materials	8.62	8.86	7.57	9.13
Demonstrates	5.14	7.06	3.69	3.65
Uses Examples	0.48	0.72	0.41	0.15
Provides Structuring Cues	6.27	9.32	6.54	3.45
Uses Directives	7.33	8.02	7.31	6.49
Questioning				
Asks Higher-Level Qs	1.57	2.25	1.79	0.53
Asks Memory Qs	5.36	5.57	7.19	3.21
Asks Opinion Qs	0.30	0.50	0.38	0.14
Asks "Do you understand?"	0.76	0.75	0.63	0.58
Probes	1.25	1.70	1.46	0.51
Responses to Questions				
Brief Response	15.68	16.17	21.71	9.18
Extended Student Response	2.03	0.91	5.29	0.36
Teacher Feedback				
Acknowledges Answer	0.90	0.74	1.39	0.35
Says "Wrong"	0.11	0.02	0.11	0.03
Repeats Answer	3.15	3.43	5.06	0.78
Gives Answer	0.22	0.20	0.11	0.13
Classroom Management				
Discipline	1.51	1.59	1.27	1.87
Procedural Interactions	7.29	10.26	5.91	10.04
Absence of Verbal Interaction	22.59	11.55	9.20	44.63

Table 2
Mean Percentages of Behaviors Recorded
(Fifth Grade Mathematics; n = 87; obs=6)

Behavior	Total	Lecture	SW/W
Teaching			
Explains	12.69	13.06	11.81
Explains with Materials	15.88	24.45	4.97
Demonstrates	3.86	4.91	2.25
Uses Examples	0.32	0.56	0.05
Provides Structuring Cues	0.83	1.23	0.34
Uses Directives	1.79	1.73	1.85
Questioning			
Asks Higher-Level Qs	0.21	0.29	0.03
Asks Memory Qs	9.81	13.43	3.89
Asks Opinion Qs	0.05	0.04	0.03
Asks "Do you understand?"	0.57	0.72	0.39
Probes	1.80	2.11	1.24
Responses to Questions			
Brief Response	12.00	16.03	5.21
Extended Student Response	2.27	1.76	1.21
Teacher Feedback			
Acknowledges Answer	1.12	1.53	0.43
Says "Wrong"	0.22	0.25	0.20
Repeats Answer	2.66	3.84	0.57
Gives Answer	0.20	0.24	0.17
Classroom Management			
Discipline	0.74	0.64	1.02
Procedural Interactions	6.33	3.96	10.18
Absence of Verbal Interaction	23.43	5.76	51.27

and gives answer). In addition, teachers in the science sample rarely acknowledged answers. Finally, teachers in the mathematics sample rarely provided structuring cues or asked higher-level questions. In general, then, a few behavioral "types" tended to occur over and over again in the observed classrooms.

Tables 3 and 4 present the intraclass correlations across all instructional formats and within each instructional format for the science and mathematics samples, respectively. All of the correlations displayed in these tables are significant beyond the 0.25 level. In addition, only correlations greater than 0.20 are displayed for the science (Table 3). For the mathematics sample correlations greater than 0.15 are displayed since these correlations are based on two fewer observations. As can be seen in these two tables, the intraclass correlations within the "total" column exceed the stated minimums for only a single behavior, demonstrates. In contrast, when the behaviors are considered within the context of the various instructional formats the intraclass correlations for several of the behaviors exceed these minimums.

Finally, Tables 5 and 6 present the estimated number of observations necessary to achieve a intraclass coefficient of 0.70 for the science and mathematics samples, respectively. When the data are considered across instructional formats, virtually all of the behaviors (with the exceptions of demonstrates and acknowledges correct answer for the science sample) would require 11 or more observations to achieve this minimum coefficient.

Again in contrast, thirteen of the twenty behaviors for the science sample would require 6 or fewer observations to achieve

Table 3

Intraclass Correlations Across All Instructional Formats and Within Each Instructional Format (Fifth Grade Science; n = 42; obs = 8)

Behavior	Total	Lect	Disc	SW/L
Teaching				
Explains				
Explains with Materials				
Demonstrates	0.20		0.41	0.28
Uses Examples		0.35	0.73	
Provides Structuring Cues				0.39
Uses Directives			0.35	0.37
Questioning				
Asks Higher-Level Qs				
Asks Memory Qs			0.48	
Asks Opinion Qs			0.82	
Asks "Do you understand"			0.23	0.21
Probes		0.33		
Responses to Questions				
Brief Response		0.30	0.39	0.37
Extended Student Response				
Teacher Feedback				
Acknowledges Correct Answer				
Says "Wrong"			0.51	0.24
Repeats Answer		0.21		
Gives Answer			0.29	
Classroom Management				
Discipline			0.59	
Procedural Interactions		0.42	0.47	
Absence of Verbal Interaction			0.39	0.35

Table 4

Intraclass Correlations Across All Instructional Formats
and Within Each Instructional Format
(Fifth Grade Mathematics; n = 87; obs = 6)

Behavior	Total	Lect	SW/W
Teaching			
Explains		0.16	0.18
Explains with Materials		0.15	
Demonstrates		0.17	
Uses Examples			
Provides Structuring Cues			
Uses Directives			
Questioning			
Asks Higher-Level Qs			
Asks Memory Qs		0.18	0.32
Asks Opinion Qs			0.24
Asks "Do you understand"			
Probes		0.17	
Responses to Questions			
Brief Response		0.23	0.32
Extended Student Response			
Teacher Feedback			
Acknowledges Correct Answer		0.19	
Says "Wrong"			
Repeats Answer			0.31
Gives Answer			
Classroom Management			
Discipline		0.22	
Procedural Interactions			
Absence of Verbal Interaction			0.21

Table 5

Number of Observation Occasions Necessary for a Generalizability Coefficient Greate than 0.70
(Fifth Grade Science; n = 42; obs = 8)

Behavior	Total	Lect	Disc	SW/L
Teaching				
Explains	11+			
Explains with Materials	11+			
Demonstrates	10		4	7
Uses Examples	11+	5	1	
Provides Structuring Cues	11+			4
Uses Directives	11+		5	4
Questioning				
Asks Higher-Level Qs	11+			
Asks Memory Qs	11+		3	
Asks Opinion Qs	11+		1	
Asks "Do you understand"	11+		8	9
Probes	11+	5		
Responses to Questions				
Brief Response	11+	6	4	5
Extended Student Response	11+			
Teacher Feedback				
Acknowledges Correct Answer	10		10	
Says "Wrong"	11+		3	
Repeats Answer	11+	9		
Gives Answer	11+		6	
Classroom Management				
Discipline	11+		2	
Procedural Interactions	11+	4	3	
Absence of Verbal Interaction	11+		4	5

Table 6

Number of Observation Occasions Necessary for a Generalizability Coefficient Greater than 0.70
(Fifth Grade Mathematics; n = 87; obs = 6)

Behavior	Total	Lect	SW/W
Teaching			
Explains	11+		
Explains with Materials	11+		
Demonstrates	11+		
Uses Examples	11+		
Provides Structuring Cues	11+		
Uses Directives	11+		
Questioning			
Asks Higher-Level Qs	11+		
Asks Memory Qs	11+		5
Asks Opinion Qs	11+		8
Asks "Do you understand"	11+		
Probes	11+		
Responses to Questions			
Brief Response	11+	8	6
Extended Student Response	11+		
Teacher Feedback			
Acknowledges Correct Answer	11+		
Says "Wrong"	11+		
Repeats Answer	11+		6
Gives Answer	11+		
Classroom Management			
Discipline	11+	9	
Procedural	11+		
Absence of Verbal Interaction	11+		9

the minimum coefficient if the instructional format within which the observations occurred was taken into consideration. For the mathematics sample only 3 of the twenty behaviors would require 6 or fewer observations to achieve the same criterion. Thus, the influence of instructional formats on the generalizability of teacher behaviors appears much stronger for the science sample than for the mathematics sample. At the same time, however, knowledge of the instructional format within which the behaviors are exhibited increases the reliability of the data beyond that possible without such knowledge in both samples.

Discussion

Two generalizations can be drawn from the results of the study. The first pertains to the concept of activity structure; the second to the nature of instruments used to observe teachers for the purpose of teacher evaluation.

The concept of activity structure appears to have potential for resolving the dilemma facing school administrators in the use of teacher observations in evaluating teachers. Considering only one of the features or dimensions of activity structures, namely, the instructional format, adequate reliability can be achieved with a somewhat more reasonable number of observations. Consideration of several of the remaining features of activity structures (particularly, pacing and cognitive level) may reduce the required number of observations even further.

In combination with earlier studies, the results of the present study suggest that the frequency of teacher behaviors can not be generalized beyond the bounds of various context factors. At the same time, however, an understanding of context factors permits one to place observers into settings and situations in which sufficiently reliable data are possible to attain.

The instrument used to gather data on the nature of teacher behaviors in the study (namely, the Five-Minute Interaction) quite clearly focused the observers' attention on verbal interactions between teachers and students. As a consequence, it is not surprising that the discourse format yielded the highest intraclass correlations and the lowest estimated number of observations to achieve a correlation of 0.70. [For the science sample slightly more than one-half of the behaviors would be associated with a intraclass correlation of at least 0.70 with six or fewer observations.] Of the four formats employed frequently enough to be included in the analysis, only the discourse format relies extensively on teacher-student verbal interaction. In fact, the non-use of the discourse format in the mathematics sample may account for the lessened effect of the instructional format on the generalizability of teacher behaviors in that sample.

The implication of this apparent "instrument-effect" is that the instrument used to observe teachers must focus on what teachers are likely to do within the instructional format or formats they are likely employ. Thus, if we know that teachers are to engage in discourse, the kinds of behaviors included on the FMI are likely be exhibited frequently. As a consequence, the reliability of the data obtained from the FMI is likely to be reasonably high under these conditions. When teachers employ the lecture format, the written seatwork format, or the laboratory seatw format, however, they are not as likely to exhibit the types of behaviors included on the FMI. Other instruments are needed to reliably observe behaviors frequently exhibited within these formats.

In combination these two generalizations support the need for additional work, both conceptually and methodologically, if sound, defensible evaluations are to be made based on observations of teachers. The concept of activity structure appears to have great promise in aiding these necessary conceptual and methodological efforts.

References

Erlich, O. and Shavelson, R. J. (1978). The search for correlations between measures of teacher behavior and student achievement: measurement problem, conceptualization problem, or both? _Journal of Educational Measurement_, _15_, 77-89.

Evertson, C. M. and Veldman, D. J. (1981). Changes over time in process measures of classroom behavior. _Journal of Educational Psychology_, _73_, 156-163.

Kowalski, J. P. S. (1978). _Evaluating teacher performance_. Arlington, Virginia: Educational Research Service.

Stallings, J. A. (1977). _Learning to look: A handbook on classroom observation and teaching models_. Belmont, California: Wadsworth Publishing Company.

Stayrook, N. and Crawford, J. (1978). _An experiment on teacher effectiveness and parent-assisted instruction in the third grade: The observational data_. Center for Educational Research at Stanford, Stanford University.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. _Educational Researcher_, _13_, 11-18.