DOCUMENT RESUME

ED 377 188 SP 035 632

AUTHOR Barnette, Jack; And Others

TITLE Evaluation of Teacher Classroom Questioning Behaviors.

INSTITUTION Appalachia Educational Lab., Charleston, W. Va.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Jul 94
CONTRACT RP91002002
NOTE 15p.; Paper presented at the Annual National Evaluation Institute (3rd, Gatlinburg, TN, July 13, 1994).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classroom Research; Elementary School Teachers; Elementary Secondary Education; Evaluation Methods; *Faculty Development; Higher Education; *Inservice Teacher Education; *Program Effectiveness; *Program Evaluation; *Questioning Techniques; Secondary School Teachers; *Skill Development
IDENTIFIERS Appalachia Educational Laboratory WV; Question Answer Reciprocity; Question Time; Question Types

ABSTRACT

This paper reports on a whole-school staff development program, field-tested and developed by the Appalachia Educational Laboratory, Inc. The program called QUILT (Questioning and Understanding To Improve Learning and Thinking) was designed to improve the questioning skills of classroom teachers by helping them individually and with colleagues. QUILT includes extensive data collection and analysis to assess effectiveness including assessment of participant knowledge, attitudes, and classroom behaviors. The behaviors of interest include: number of teacher and student initiated questions, use of wait-time, cognitive levels of questions and student answers, manner of designating students to answer questions, and use of various types of desirable and undesirable teacher responses or feedback. The aspect of QUILT examined here, the evaluation of teacher classroom questioning behaviors, is based on classroom observation, pre- and post-QUILT assessment of teacher knowledge, attitudes, classroom questioning practices, videotaped observation and coding of teacher behaviors. Data analysis revealed strong evidence that teacher behavior is positively influenced by the QUILT program. (Contains 19 references.) (LL)

SP

Evaluation of Teacher Classroom Questioning Behaviors

by

Jack Barnette, University of Alabama
and
Sandra Orletsky and Beth Sattes,
Appalachia Educational Laboratory

Paper presented at The Third Annual National Evaluation Institute, Gatlinburg, TN, Field Symposium, July 13, 1994.

This paper is based on work sponsored wholly or in part by the Office of Educational Research and Improvement, U.S. Department of Education, under contract number RP91002002. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

The teacher evaluation activity reported here is an integral part of a staff development program developed by the Appalachia Educational Laboratory, Inc. (AEL) called QUILT, designed to improve classroom teacher questioning skills. QUILT includes extensive data collection and analysis to assess effectiveness including assessment of participant knowledge, attitudes, and classroom behaviors. The aspect of QUILT research reported here is the evaluation of teacher classroom questioning behaviors based on classroom observation.

Research indicates that as much as 40% of classroom time is spent in a question-response mode (Johnson, Markle, & Haley-Olpihant, 1987). Nevertheless, many teachers do not ask questions effectively (Gall, 1984). Ineffective or inappropriate practices include asking questions at only lower cognitive levels (Ornstein, 1987), directing a disproportionate percentage of questions toward a limited number of students (Jones, 1990), or waiting too little time after asking a question before reacting to the student response, typically one second or less (Rowe, 1986). Questions too often flow in only one direction and become a way of maintaining control rather than stimulating thought. For example, teachers are likely to ask as many as 50 questions during a typical class period while it is unlikely that the students in the class ask even one question (McGlathery, 1978).

QUILT Overview

QUILT is a staff development program for classroom teachers. Its goal is to provide a focal point for whole-school staff development by helping teachers, individually and with colleagues to improve their skills in asking questions, a teaching strategy used in all subjects K-12. QUILT stands for Questioning and Understanding to Improve Learning and Thinking. It was designed and field tested by the Appalachia Educational Laboratory (AEL) of Charleston, WV, in collaboration with school personnel in Kentucky, North Carolina, Tennessee, Virginia, and West Virginia.

QUILT has four components: induction training, collegiums, partnering, and independent study and analysis. Induction training is a three-day, 18-hour program, conducted by trainers trained by AEL, where participants are provided research-based knowledge and theory, as well as frequent opportunities to practice effective questioning techniques. Experience indicates that the three-day induction period produces a degree of bonding among participants that does not occur during shorter sessions. The selection of QUILT as the program acronym had a major affect on development of participant bonding. Not only have the program components and training been designed around the development of a quilt, but the stories about family quilts shared by participants in getting acquainted sessions have resulted in within group personal cohesiveness rarely seen in staff development.

During the school year, teachers and administrators meet seven

times in forums designed to review information about questioning and reinforce changes in teacher questioning behaviors. These are referred to as collegiums and, although they are open-ended, each of these focuses on greater understanding and reinforcing particular questioning skills and behaviors. Partnering involves teams of peer teachers in ongoing, mutual support activities within the schools. These activities include visiting each other's classrooms to observe and monitor progress in questioning and to provide support and encouragement. Throughout the year participants read independently, practice their skills, and compile data on their own classroom behaviors and student responses.

QUILT differs in significant ways from the approaches to staff development most frequently employed by schools. First, QUILT treats staff development as a long term commitment. Research indicates that only a small percentage of teachers, perhaps as low as ten percent, change their behavior in response to a training program unless lectures or seminars are reinforced by feedback in a classroom setting (Joyce & Showers, 1982). QUILT is a multi-year program and the "partnering" approach to classroom instruction is central to its design.

Second, QUILT represents a "whole-school" approach to staff development. Because questioning is a generic educational activity, improving questioning skills is relevant across curriculum from kindergarten to the 12th grade. The partnering approach reflects this generic quality since teachers across subject areas can work together to improve questioning skills.

Third, QUILT is student-centered. While it is fashionable to make this claim for almost any program, the entire purpose of the QUILT five-stage model is to stimulate student thinking, particularly higher order thinking. As Dillon (1984) states, "to conceive an educative question requires thought; to formulate it requires labor; and to pose it, tact." The QUILT five-stage model helps teachers view questioning as a process which begins with planning the question and ends with reflectively evaluating the effectiveness of the questioning episode. More specifically, the five stages are as follows. Stage 1 relates to preparing the question. It includes identifying the instructional purpose, determining the content focus, selecting the appropriate cognitive level, and considering wording and context. Stage 2 relates to presenting the question. It includes indicating the response format, asking the question, and respondent selection. Stage 3 relates to prompting student response. It includes pausing after asking the question, assisting nonrespondents, and pausing after student response. Stage 4 relates to processing the student response. It includes provision of appropriate feedback, expanding and using correct responses, and eliciting student reactions and questions. Stage 5 relates to critiquing the questioning episode. It relates to analyzing the question, mapping respondent selection, evaluating student response patterns, and examining teacher and student reactions.

During the field test year, QUILT was implemented in a manner

which permitted assessment of its effectiveness. Three levels of implementation of QUILT components were initiated. Schools were randomly assigned into one of the three groups. Group A schools completed the full QUILT program which included three-day induction, collegiums, partnering, and other independent study activities. Group B schools completed only the three-day induction program and Group C schools received only a three-hour orientation session related to QUILT questioning concepts.

An extensive research design was developed to assess QUILT effectiveness (Barnette & Sattes, 1991). This included pre and post-QUILT assessment of teacher knowledge, attitudes, and classroom questioning practices. In addition, evaluation of all aspects of program delivery and implementation was conducted. The aspect of QUILT research reported here is related to the evaluation of teacher classroom questioning practices based on the videotaped observation and coding of teacher behaviors.

Development of the Classroom Questioning Observation Instrument

The Classroom Questioning Observation Instrument (CQOI) was developed as one data-gathering tool needed for the QUILT research design. Its primary purpose is to collect specific information on teachers' classroom questioning behaviors with the data being used to help analyze teacher classroom questioning behavior change. Specifically, it was designed to address one of the QUILT research hypotheses (Barnette & Sattes, 1991), namely:

> There will be a significant difference between the three groups on the dependent variables related to teacher classroom questioning behavior, as measured by the Classroom Questioning Observation Instrument. These differences will be directional in nature, with Condition A having the highest level of desirable behaviors, Condition B having the second highest, and Condition C having the lowest level of desirable behaviors.

More specifically, the behaviors of interest included: number of teacher and student initiated questions, use of wait-time I, use of wait-time II, cognitive levels of questions and student answers, manner of designating students to answer questions, and use of various types of desirable and undesirable teacher responses or feedback.

Because participating teachers were spread out over five states and in an attempt to reduce obtrusiveness of an actual observer in the classroom, it was decided to have 15-minute videotapes recorded, which would be reviewed and coded by trained coders. The CQOI is a low inference, multiple code, category system observation instrument. As such, it was designed and developed with these factors in mind:

1. a format which provided for ease of data collection and analysis,

2. clearly stated definitions to increase coder reliability and inter-rater agreement,

3. a direct connection to the QUILT materials, and

4. the desired outcomes, both in terms of research and the usability of information to teachers participating in the QUILT program.

Dr. Debra Sullivan, the CQOI developer, used prior knowledge of other classroom observation instruments, QUILT materials, and classroom visits to design the instrument. Throughout the instrument's formative stages of development, the developer visited classrooms and collected data using draft versions of the instrument. Using this process, not only was it possible to assess specific research questions, but "real life" usability in classroom situations was assured. Meetings were held with AEL staff to ensure a match between the research design and the teacher behavior data collection device. Points raised at these meetings were used to modify the CQOI, increasing the level of content validity.

For logistic reasons, it was decided to have all coders living in the Charleston area. Four middle school and high school teachers were selected by the CQOI developer to participate in coder training. All of the selected teachers were considered extremely capable and competent teachers who represented several major curricular areas including language arts, social studies, mathematics, science, and foreign language.

Coders were trained using a variety of methods including group sessions as well as independent work. During the training sessions, coders:

1. were acquainted with the QUILT program and its research design,

2. were familiarized with the CQOI in terms of format, definitions, and manner of completion,

3. practiced coding transcripts of classroom sequences featuring questioning interactions between teachers and students, and

4. practiced coding videotapes of classroom episodes.

Similarly, the independent work completed by coders focused on written transcripts of classroom interactions as well as classroom videotapes. During the coder training, CQCI codes and their definitions were discussed and defined more clearly, thus ultimately assuring higher levels of coder validity and reliability.

Since 15-minute videotapes of classroom teaching episodes were used rather than direct observation, coder speed was not an area of concern. Coders were able to replay the tape to check coding for

accuracy and reconsideration. Therefore, only accuracy in coding classroom questioning behaviors was necessary to determine coder reliability. Reliability was established by comparing coder responses with those of the CQOI developer on the same videotape. The range of agreement of coding ranged from 90 to 94%, with an average agreement of 92%. Coders did not know the teachers who were observed, nor did they know which QUILT condition they represented.

The CQOI permits the, mostly linear, coding of many teacher behaviors and characteristics of questioning in the classroom. Each questioning episode is recorded in terms of whether it was teacher or student initiated. For teacher initiated questions, whether the teacher designated a student to answer before or after asking the question is then recorded. The level of question is recorded as being recall, check for understanding, utilization, or creation. Wait-time I, the time a teacher waits before acknowledging a student response to an initial question, is recorded by checking the number of seconds. The student answering, whether the one designated before or after the question was asked, is recorded. The number of students responding is recorded as one, more than one, or whole class (choral response). The level of student answer is recorded as being recall, check for understanding, utilization, or creation. The student answer is also recorded as being correct, partially correct, wrong, no answer, inappropriate response and if the student asks for clarification or extends his/her answer.

Wait-time II is recorded, the time the teacher waits before reacting to the student answer. The teacher reaction is recorded as being positive feedback, praise, negative feedback, corrective feedback, criticism, or no feedback given. In addition, other teacher behaviors are recorded including whether the teacher probes, repeats or rephrases the question, repeats or rephrases the student answer, uses the student response in discussion or new questions, and/or redirects the question to other students.

Data Analysis

It is beyond the scope of this paper to present all of the results of the comparison of pre-QUILT and post-QUILT classroom behaviors across the three QUILT training conditions. This analysis is limited to seven of the more important variables measured by coding of the observations prior to the start of QUILT (referred to as pre) and again at the end of the first complete year of QUILT operation (referred to as post). Three different groups were included in this analysis:

Condition A (Full QUILT model including induction and collegiums)

Condition B (QUILT induction without collegiums)

Condition C (QUILT awareness workshop)

The data were analyzed using several programs of the SAS package (SAS is a registered trademark of SAS Institute Inc., Cary, NC). For each QUILT variable, the following analyses were conducted:

1. Univariate summary statistics were computed for pre test results, post test results, and post-pre test results. Included were tests for normality and provision of data for computation of F_{max} statistics for checking analysis of variance assumptions. These results were used to compute effect sizes. The pre test standard deviation for participant scores in all three groups was used as the base for the effect size. The post test minus pre test means were divided by the overall pretest standard deviation to obtain the effect size.

2. The GLM procedure was conducted as a mixed design, with a between subjects factor (condition) and a within subjects factor (testing time). Of primary concern were two planned follow-ups of the interaction. Since these comparisons were in the planned mode, the significant interaction of condition and time was not required to conduct these follow-ups.

3. The first follow-up procedure involved the comparison of pre and post test means within each condition. These were compared using directional, dependent t tests with alpha set at 0.05.

4. The second follow-up procedure involved the comparison of post test means of condition A with each of the other groups (A with B and A with C). These were compared using directional, Dunnett t tests with alpha set at 0.05. The Dunnett is specifically designed to compare groups with a control group or the situation where all groups are compared with only one other group. In this case condition A was compared with each of the other groups. Dunnett controls Type I error rate in an experiment-wise manner. It is one of the few planned follow-up procedures which can be used to test directional hypotheses. Thus, it has high statistical power, but is limited to the number of groups, minus one, pairwise comparisons.

5. The third follow-up procedure involved the comparison of the pre to post test change mean of condition A with each of the other groups (A with B and A with C). These were compared using directional, Dunnett t tests with alpha set at 0.05.

Results

Number of Teacher Questions

During the 15 minute video tape, the number of teacher initiated questions was recorded. The desirable change was that there be a decrease on this variable. Results for this variable are presented in

Table 1. All three groups had reductions in the number of teacher questions. This reduction was significant for conditions A and C, with effect sizes of -0.65 for condition A and -0.44 for condition C.

Table 1. Number of Teacher Questions by QUILT Condition

QUILT Group	Mean/SD Pre	Mean/SD Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	41.4	31.0	-0.65	<0.001	A<B	none
n= 37	15.8	14.5				
Condition B	44.9	40.5	-0.27	nsd		
n= 28	17.4	13.8				
Condition C	43.3	36.3	-0.44	<0.05		
n= 30	15.5	14.4				

At post, condition A had a significantly lower number of teacher initiated questions than condition B. There were no significant differences between condition A and conditions B or C relative to the degree of change between pre and post test, although the difference was in the predicted direction.

Wait-Time I

Wait-time I is the time a teacher waits after asking a question before acknowledging or reacting to a students response. It is recommended that this time be three seconds or longer. Results for

Table 2. Wait-time I, Percentage at Three or More Seconds by QUILT Condition

QUILT Group	Mean/SD Pre	Mean/SD Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	12.8	25.0	+0.99	<0.01	A>C	A>C
n= 37	11.9	24.9				
Condition B	11.1	20.7	+0.78	<0.01		
n= 28	10.1	19.5				
Condition C	10.1	11.5	+0.11	nsd		
n= 30	14.8	16.5				

this variable are found in Table 2. It was predicted that this variable would increase as a result of QUILT. Both conditions A and B had significant increases in this variable, with effect sizes of +0.99 for condition A and +0.78 for condition B. Condition A had a higher mean at post as well as significantly more pre to post change as compared with condition C.

Wait-time II

Wait-time II is the time a teacher waits after a student's response to a question before acknowledging or reacting to that response. It is recommended that this time be three seconds or longer. Results for this variable are found in Table 3. It was desired that

Table 3. Wait-time II, Percentage at Three or More Seconds by QUILT Condition

QUILT Group	Mean/SD Pre	Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	0.52	2.98	+1.72	<0.05	A>B	A>C
n= 37	1.28	6.73			A>C	
Condition B	0.10	0.59	+0.34	nsd		
n= 28	0.51	1.61				
Condition C	0.59	0.97	+0.26	nsd		
n= 30	2.06	4.57				

this variable increase. While the level of the use of wait-time II is very low, there was a significant pre to post change for condition A, effect size of +1.72. At post, condition A was significantly higher than both conditions B and C. Condition A had a significantly higher pre to post change as compared with condition C.

Teacher Questions Above Recall Levels

The percent of times teacher initiated questions that were above recall cognitive level was determined and results presented in Table 4. An objective of QUILT training is to increase the frequency of higher level questioning. Condition A was the only group to have a significant pre to post change, with an effect size of +0.43.

Table 4. Cognitive Level of Question, Percentage above Recall Level, by QUILT Condition

QUILT Group	Mean/SD Pre	Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	31.0	41.2	+0.43	<0.05	none	none
n= 37	23.3	27.8				
Condition B	41.0	39.2	-0.07	nsd		
n= 28	24.8	30.1				
Condition C	26.3	32.0	+0.24	nsd		
n= 30	22.0	22.7				

Percentage of Time Teacher Redirects Question to Other Student(s)

The percentage of times a teacher redirects a question to other student(s) was determined and is reported in Table 5. A QUILT objective was for this to increase. Condition A had a significant pre

Table 5. Question Redirected to Another Student Percentage by QUILT Condition

QUILT Group	Mean/SD Pre	Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	14.1	23.2	+0.59	<0.01	A>C	A>B
n= 37	14.5	19.9				A>C
Condition B	20.6	19.4	-0.08	nsd		
n= 28	16.7	14.9				
Condition C	18.1	12.3	-0.37	nsd		
n= 30	15.0	14.5				

post change with an effect size of +0.59. At post, condition A had a significantly higher mean than condition C and condition A had significantly higher pre to post change than both conditions B and C.

Percentage of Time Student Designated to Answer After Question Asked

The percent of times the teacher designated which student was to answer a question after it was asked was determined and results

presented in Table 6. It was a QUILT objective to increase this practice because often when the student is designated prior to the question rather than after the question, other students, since they feel they are not involved, reduce their level of involvement or discontinued involvement in the interaction totally. Both

Table 6. Student Designated after Question Percentage by QUILT Condition

QUILT Group	Mean/SD Pre	Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	84.1	90.8	+0.39	<0.01	A>B	none
n= 37	12.8	9.3				
Condition B	83.1	85.3	+0.13	nsd		
n= 28	23.2	11.2				
Condition C	83.5	89.4	+0.35	<0.05		
n= 30	14.4	11.2				

conditions A and C had significant pre to post changes, with the effect size for condition A at +0.39 and for condition C at +0.35. At post, the condition A mean was significantly higher than condition B.

Percentage of Time Teacher Repeats Student Answer

Another variable, which QUILT was designed to decrease was the percentage of time a teacher repeats the student answer. Often when

Table 7. Teacher Repeats Student Answer Percentage by QUILT Condition

QUILT Group	Mean/SD Pre	Post	ES	p	Significant Group Differences with Condition A at Post	Change
Condition A	62.4	54.6	-0.43	<0.05	none	none
n= 37	18.9	28.5				
Condition B	60.5	55.9	-0.25	nsd		
n= 28	14.3	17.9				
Condition C	59.4	61.5	+0.11	nsd		
n= 30	20.9	25.5				

this happens other students take this as acknowledging the response as being correct and then there is no need to continue thinking. If the teacher gets the answer, students often tune-out. Table 7 presents the results for this variable. Condition A was the only one to have a significant pre to post reduction in this behavior, with an effect size of -0.43.

Summary of Differences Between Condition A and Other Groups at Post

Condition A had a significantly lower number of teacher questions at post than condition B. It had a significantly higher percentage of wait-time I of three seconds or higher as compared with condition C. Condition A had significantly higher percentage of wait-time II at three seconds or more than either of the other two conditions. Condition A had a significantly higher percentage of times the teacher redirects the question to other student(s) as compared with condition C. Condition A had a significantly higher percentage of times the teacher designated the student to answer after asking the question as compared with condition B.

Summary of Differences Between Pre-Post Change of Condition A and Other Groups

Condition A had significantly more desirable pre to post change than condition C on the variables of wait-time I, wait-time II, and question redirection to other student(s). Condition A had significantly more desirable pre to post change than condition B on the variable of question redirection to other student(s).

Conclusions

It is clear that condition A had the greatest degree of change in predicted and/or desirable directions. On all seven of the selected variables of the CQOI, there was a significant pre to post change for condition A, compared with one change for condition B, and two such changes for condition C. On post comparisons, the condition A mean was more favorable than the condition B mean on three of the variables and more favorable than the condition C mean on three of the variables. Condition A had higher positive pre to post change than condition B on one of the variables and higher than condition C on three of the variables. Since these teachers were randomly selected and assigned to the three treatment conditions, there is strong evidence that teacher behavior has been positively influenced by the QUILT program. The use of observational data provides information helpful in answering the questions related to the impact of staff development in terms of going beyond just learning more about classroom questioning, but being able to apply learning to the actual classroom situation. The CQOI provided a reliable method of evaluating teacher classroom questioning behaviors yet was relatively unobtrusive and valid.

References

Adams, R. & Biddle, B. (1970). Realities of teaching. New York: Holt, Rinehart & Winston.

Adler, M. J. (1984). The Paideia program. New York: Macmillan.

Barnette, J. J. & Sattes, B. D. (1991). QUILT data collection coordinator's manual. Charleston, WV: Appalachia Educational Laboratory.

Barnette, J. J., Sullivan, D., Orletsky, S. R., & Sattes, B. D. (1993). Observation of teacher classroom questioning behaviors. Paper presented at the 1993 Annual Meeting of the American Educational Research Association, Atlanta, Ga.

Bloom, B. S., (Ed.). (1956). Taxonomy of educational objectives: Cognitive domain. New York: Longman.

Brophy, J. (1981). Teacher praise: A functional analysis. Review of Educational Research, 51(1), 5-32.

Brophy, J. & Good, T. (1980). Educational psychology, A realistic approach. New York: Holt, Rinehart & Winston.

Dillon, J. T. (1983). Teaching and the art of questioning. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Dillon, J. T. (1984). Research on questioning and discussion. Educational Leadership, 42(3), 50-56.

Dillon, J. T. (1988). Questioning and teaching: A manual of practice. New York: Teachers College Press.

Gall, M. (1984). Synthesis of research on teachers' questioning. Educational Leadership, 42(3), 40-47.

Hunkins, F. P. (1989). Teaching thinking through effective questioning. Boston: Christopher-Gordon Publishers, Inc.

Johnson, J. H., Markle, G. C., & Haley-Oliphant, A. (1987). About questioning in the classroom. Middle School Journal, 18(4), 29-33.

Jones, M. G. (1990). T-zone, target students and science classroom interactions. Journal for Research in Science Teaching, 27(7), 631-660.

Joyce, R. B. & Showers, B. (1982). The coaching of teaching. Educational Leadership, 40(1), 4-10.

McGlathery, G. (1978). Analyzing the questioning behaviors of science teachers. In M. B. Rowe (Ed.), What research says to the science teacher, Volume 1. Washington, DC: National Science Teachers Association.

Ornstein, A. C. (1987). Questioning: The essence of good teaching. NASSP Bulletin, 71(499), 71-79.

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up. Journal of Teacher Education, 37(1), 43-48.

Walsh, J. A. & Sattes, B. D. (1991). QUILT induction manual. Charleston, WV: Appalachia Educational Laboratory.