DOCUMENT RESUME

ED 330 711 TM 016 281

AUTHOR McNeil, Keith; Blanchard, Charles
TITLE How Program Staff Can Provide False Evaluation Data: The Project SEED Experience.
PUB DATE Oct 90
NOTE 11p.; Paper presented at the Annual Meeting of the Midwestern Educational Research Association (12th, Chicago, IL, October 17-19, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bias; *Data Collection; Data Interpretation; Elementary Secondary Education; *Evaluation Problems; Evaluation Utilization; Evaluators; Graduate Students; Higher Education; *Interviews; Mathematics Instruction; *Program Evaluation; Program Implementation; Staff Role; Surveys
IDENTIFIERS *False Evaluation Data; *Project SEED; Stakeholder Evaluation

ABSTRACT

Evaluators usually assume that the data they receive are valid, or at least, that the data reflect that which is being evaluated. In a recent evaluation in a large southern school district, evaluators compared their raw interview notes with interview notes edited by the teachers interviewed or project staff. The program evaluated was Project SEED, an additional mathematics class for students of low socioeconomic status. In the present study, 24 graduate students with no knowledge of the background compared 14 original and edited pairs of comments/notes in terms of whether the edited notes: (1) contained additional information; (2) were more positive in tone and intent; (3) were less concrete; and (4) changed meaning or content. Edited responses contained additional information and were more positive about the project, in the opinion of naive raters. Recommendations are made to control the collection of data so that the evaluation will be useful to decision makers. Allowing those who are being evaluated to control and possibly modify the data raises the danger of bias. The raw and edited responses, the survey form administered to the graduate student raters, and a table of the results are provided. (SLD)

HOW PROGRAM STAFF CAN PROVIDE FALSE EVALUATION DATA: THE PROJECT SEED EXPERIENCE

Keith McNeil and Charles Blanchard
New Mexico State University

HOW PROGRAM STAFF CAN PROVIDE FALSE EVALUATION DATA: THE PROJECT SEED EXPERIENCE

Keith McNeil and Charles Blanchard New Mexico State University

Objective: The paper will empirically demonstrate how one project staff provided false evaluation data. Also included are several hints as to how evaluators could avoid this type of situation in the future.

Perspective: Evaluators usually assume that the data they receive are valid, or at least that the data reflect that being evaluated. In a recent evaluation in a large Southern school district evaluators had an opportunity to compare "raw interview notes" with "edited interview notes." The raw interview notes were edited by project staff and were judged by the evaluators to be so very different from the raw notes that the interview data was basically not included in the final report. The present study used raters who had no knowledge of the situation and hence had no vested interest in the outcome to determine if the edited responses were indeed different.

Techniques: The project being evaluated was called "Project SEED," a 45 minute per day additional mathematics class for low SES students. The curriculum of SEED consisted primarily of logarithms and exponents and was presented by "mathematics experts," few of whom had teaching degrees. Consequently, the regular teachers had to remain in the classroom, performing only perfunctory classroom duties during that 45-minute period. Thus was the program not only costly, but focused on content that was neither on the district-wide test nor on the list of state competencies. Never the less, the project director had convinced the school board and the Superintendent that SEED was a good program. Indeed, at the beginning of the evaluation, the Superintendent distributed a memo to his staff wherein he stated that the SEED project was the "best program in my 29 years of educational experience" (Superintendent memo in McNeil and Blanchard, 1989).

Because the project was being implemented in six schools that were under a court desegregation order, the judge ordered that the project be evaluated. The project director's numerous attempts to control or destroy the evaluation have been presented elsewhere (McNeil and Blanchard,1989). The present paper focuses on one aspect of the evaluation--interviews of the SEED teachers.

Method: A plan for interviewing the teachers was prepared and was shared with the project director. The plan specified a random sample, but the project director wanted to construct his own

---

Paper presented at the Annual Meeting of the
Midwestern Educational Research Association
October 17-20, 1990, Chicago, Illinois

---

sample, and balked at our plans to sample by saying that we might obtain in our sample only the new teachers, those with the least experience, etc. He was requested to provide the relevant variables and the scores of each teacher on those variables so the sample could be stratified on those variables. He refused, and pulled his political strings such that his "sample" was the one that was finally used.

Two evaluators (the first author and his boss) interviewed three teachers each with the structured interview. The interviewers wrote down the responses and read them back to the teachers to make sure that they had accurately captured the essence of the response. Once the notes were typed and proofed by the evaluators, the typed notes were transmitted to the project director with the request that the notes be edited by the teachers. Edited they were, as indicated in Appendix A. The evaluators received what we described as "highly edited" notes. The original notes were not returned (fortunately copies had been kept). Each of the teacher's notes were totally retyped with the same machine. The evaluator's conception of editing for clarification or grammar was implemented by the project staff in what can only be descried as a "snow job." In many cases the edited comments were judged by the evaluators to be totally unrelated to the original comments. In other cases, the edited comments were seen as depicting the project in a much more favorable light. Since the evaluators felt that the project staff (actually most likely the Project Director) did not cooperate with the intent of the interview effort, the evaluators did not feel obliged to report these results in detail. Consequently, the entire effort was summarized in the final report as: "The six SEED teachers identified by SEED management were extremely positive about SEED and their role in the project. In terms of the students of the project, SEED teachers maintained that students gained not only additional math knowledge, but also increased self esteem. The major weakness identified by the SEED teachers was the constraint of each student taking only one semester each year." (McNeil, 1987)

Rating Data: In the present study raters who were unaware of the background situation, rated 14 pairs of comments. Out of the 30 comments (6 SEED teachers times 5 questions), 16 comments were omitted for one of the following two reasons: 1) comments were so different that they couldn't be compared, or 2) comments required too much knowledge of the project or setting.

Each rater rated the original and edited pairs of comments on each of four scales. (See Appendix B for an example of the rating form.) These scales were chosen because they were suspected to be the major ways that the edited responses differed from the original:

1. contains additional information.
2. more positive in tone and intent.
3. less concrete.
4. changes meaning or intent.

Results: The ratings were obtained from 24 graduate students who were unaware of the background of the project and its attempts to avoid being evaluated. The apriori rating differences between the original and edited versions were tested with the one-tailed correlated t test, because of the apriori hypotheses. Table 1 contains the results, summed across the 14 items, as well as summary information for the individual 14 responses.

Question 1. Our hypothesis that the edited ratings contained additional information was supported in 5 out of the 14 comparisons. When the results were summed over the 14 items, the hypothesis was also supported. Thus naive raters felt that the edited responses did contain additional content, as compared to the original responses.

Question 2. Our hypothesis that the edited ratings were more positive in their tone and intent was supported in 7 out of the 14 comparisons. When the results were summed over the 14 comparisons, the hypothesis was also supported. Thus, our initial and possibly emotionally biased reaction that the edited responses cast the SEED project in a more positive light than did the raw responses was upheld by the naive raters.

Question 3. Our hypothesis that the edited ratings were less concrete was not supported. In looking at the results for each of the 14 comparisons, we noted that question 3 tended to be rated in much the same way as the first two questions. While the raters were cautioned to read each comparison and each question carefully, we have no assurance that the raters carefully responded to the task. Questions 1 and 2 may well have set up a response bias in how questions 3 and 4 were rated.

Question 4. Our hypothesis that the edited ratings changed the meaning or intent was supported in 2 of the 14 comparisons. With hindsight, this was not a very good question to ask naive raters, because the raw and edited responses did, in a general sense, provide similar information. Only someone familiar with the entire situation could realize the ramifications of some of the small changes. Furthermore, we eliminated from the rating form those comparisons that changed the meaning so much that we felt raters could not make meaningful ratings.

What was really different in the raw and edited ratings was the subtle message that the project was successful and that there were no problems within the project. Not only did the edited responses contain additional information, as judged by naive raters, but these raters also believed that the edited responses were more positive about the SEED project.

Recommendations: Due to the experiences of evaluating SEED, and the supporting evidence provided by the raters, the following recommendations are made:

1. Do not allow the data gathering process out of sight of the evaluators.
2. Make clear to those being evaluated why certain evaluation process are being followed. When projects resist being evaluated, try to explain the benefits of being evaluated and the consequences

of not being evaluated.
3. Make sure that the evaluation has the support all the way up the administrative ladder. In this case, even though the project was ordered by a Federal court to be evaluated, the project director had the ear of the superintendent and successfully fought the evaluation all the way.
4. Report any tampering of data when it occurs; hopefully the organization will deal with the issue.
5. If the organization cannot deal with the issue, or refuses to deal with the issue, then there should be a central clearing place to report these findings, as discussed by Slavin (1990).
6. If all avenues are blocked, then live with the problem until you get out of the organization, and then write a conference paper (or two depending upon the circumstances).

Conclusion: In order for evaluation data to be useful to decision makers, the data must be valid. Allowing those who are being evaluated to take control of the data and possibly modify the data may result in the data becoming biased in a more positive direction. If data are suspected of being manipulated, then the evaluator has the right, indeed the obligation, to either report the tampering or to at least not report those data.

Epilogue:

1. A five person team was evaluating the entire operation of six schools. The two persons who had responsibility for evaluating the SEED project were removed from that evaluation team for the subsequent year.

2. SEED has not been evaluated in that District in the succeeding three years.

3. SEED was given a five-year contract to continue operations in the District during the last Board Meeting that the Superintendent attended, before he left the District.

4. The SEED budget and effort has increased as the curriculum is now in some of the middle schools.

5. SEED was also asked to explore implementing their techniques in the Language Arts curriculum.

6. The leader of the evaluation team is no longer in the District.

REFERENCES

McNeil, K. (1987,July). Evaluation of the South and West Dallas Learning Centers Final Report, 1986-87 (REA87-015-10). Dallas, Texas: Dallas Independent School District, Department of Research, Evaluation and Audit.

McNeil, K. & Blanchard, C. (1989, November). An "evaluation" of a commercial curriculum. Paper presented at the meeting of the Rocky Mountain Educational Research Association, Tulsa, OK.

Slavin, R. (1990). On making a difference. Educational Researcher 30-34 April.

APPENDIX A: THE 14 RAW AND EDITED RESPONSES

RAW RESPONSES

1. Because teaching it all 3 years can have a different curriculum over the three years, leading to more sophisticated curriculum.
   SEED teachers know the regular teachers real well because they have been there two previous years for half a semester.
2. Positive Impact - teaches habits of success throughout life
   Critical thinking skills
   Lessen the fears students have to articulate an opinion
3. The West Dallas Centers were prepared for the project.
   West Dallas center students need their self image improved
   The teaching methodology that is different is an advantage
4. Not enough money for increased salaries, to put some stability (e.g., retirement). Too much driving.
5. Assuming there are weaknesses, we can't spend more time with students. Because some students are only taught for a semester.
6. Increased funding to allow SEED classes to go the entire year would be desirable.
7. Scheduling could improve driving problems, but probably not since they had all classes to choose from the first semester.
8. The option to continue with some classes as suggested by teacher.
   Recognizing we are not perfect, but continue to improve.
9. Have more than 45 minutes with each class.
10 math content--Very effective--curriculum is geared to see the big picture of mathematics; see connections between all these different elements. Deeper understanding of fractions.
11 Would want students to generalize the preciseness (i.e., above, on top of, read entire problem and look for operation instead of automatically adding).
12. Hope they develop a level of confidence to ask questions about other subjects.
    Want to participate and become involved in other subjects.
    They will know that they can.
13 If they feel good about themselves they'll want to do well.
   Take fear away from asking questions. Desensitize students to questions.
14 Their ability to speak. Curriculum - goes from one-word answers to why.
   Students will be able to give a clear response and feel good about learning.

EDITED RESPONSES

1. A three semester exposure allows students to discuss sophisticated topics in mathematics. The partitioning of our curriculum. Assigning different topics to each grade, allows students to review material in an exciting new context.
   The SEED specialists know the regular teachers from observing classes in previous years. This helps foster better communication between the regular teacher and the SEED

specialist.

2. Project SEED teaches habits of success, thereby giving them confidence they need to carry them through high school and college and into careers worthy of their ability. Project SEED's teaching methods improve students' critical thinking skills. These methods also lessen the fear students may have to articulate.
3. The professionalism of the Project SEED mathematicians
   The group discovery (Socratic) methodology used.
   The high level of the mathematics which is taught and continually tailored to the needs of the class.
   The above combination (mathematicians, discovery teaching, high-level mathematics) provides the basis for the enhancement of the student's self-image.
4. Weaknesses of Project SEED in the Learning Centers, like those elsewhere, arise basically from shortage of money: low salaries and not enough sense of stability (e.g., no retirement), long hours and too much driving
5. Students are limited to only one to only one semester of Project SEED per academic year.
6. Increased funding to allow SEED to go the entire year.
7. Conditions would improve if one could schedule classes to cut down driving. Since, however, in the second semester we have to teach exactly those classes which were not taught in the fall, scheduling will be, if anything, more difficult.
8. Occasionally, a teacher will request an extension of the SEED class which cannot be granted due to scheduling constraints. Overall, the effectiveness of the SEED program is improved by regularly resisting stagnation in the midst of an on-going quest for excellence.
9. Students could be exposed to SEED more times per class period or more class periods per year.
10. <u>math content</u>--Very effective--curriculum is geared to show the big picture of mathematics; it illustrates connections between many different elements. A deeper understanding of fractions is one important benefit.
11. I would want students to extend their inquisitiveness and curiosity to other subjects; not to be afraid of advancing and defending their own ideas and to use precision in expressing themselves.
12. I would hope that as a result of the SEED experience, students would have acquired an increased hunger for knowledge and the courage and perseverance to actively seek such knowledge, regardless of any risks or obstacles that might otherwise hinder this pursuit.
13. If they feel good about themselves, they will want to do well in all subjects.
    We desensitize students to questions. We reduce the anxiety associated with answering questions.
14. The objective would be for the students to not only speak precisely and cogently, as a result of thinking clearly and logically, but to feel good about acquiring such skills.

APPENDIX B RATING FORM

SEED RATINGS

A special math project exists in Dallas Texas, called "Project SEED." The project provides 45 minutes of supplementary math instruction to students in grades 4, 5, and 6 in three schools in West Dallas and three schools in South Dallas. These schools are in low socioeconomic areas of the city. The teachers in the project are not regular employees of the district, but are full-time employees of the company that has the contract to conduct Project SEED. SEED is an acronym for Supplementary Early Education for the Disadvantaged.

Survey responses were obtained from the same teachers on two different occasions. We need your help in comparing these responses. Please read the survey question and the pair of responses. The statement on the left and the statement on the right are the two responses to that question by one teacher. You are to compare these response on the basis of the four attributes that are at the bottom of the page. Please circle the rating which most reflects your reaction to the statement comparison.

We thank you for participating and will provide the result of this study back to you.

Please read the following pair of responses. The statement on the left and the statement on the right are two responses from the same teacher. You are to compare these responses on the basis of the four questions below. Please circle your reaction to the statement comparison.

WHAT ARE THE WEAKNESSES OF PROJECT SEED IN THE LEARNING CENTERS?

((EDITED))	((RAW))
8. Occasionally, a teacher will request an extension of the SEED class which cannot be granted due to scheduling constraints. Overall, the effectiveness of the SEED program is supported by regularly resisting stagnation in the midst of an on-going quest for excellence.	The option to continue with some classes as suggested by teacher. recognizing that we are not perfect, but continue to improve.

1. Compared to the statement on the <u>left</u>, the statement on the <u>right</u> contains additional information.

Strongly Disagree Disagree Neutral Agree Strongly Agree

2. Compared to the statement on the <u>left</u>, the statement on the <u>right</u> is more positive in it's tone and intent.

Strongly Disagree Disagree Neutral Agree Strongly Agree

3. The statement on the <u>right</u> is more concrete than the statement on the <u>left</u>.

Strongly Disagree Disagree Neutral Agree Strongly Agree

4. The statement on the <u>right</u> changes the meaning or intent of the statement on the <u>left</u>.

Strongly Disagree Disagree Neutral Agree Strongly Agree

TABLE 1. RESULTS FOR THE FOUR RATINGS.

Question	Significant Comparisons	Probability
1. Compared to the statement on the left, the statement on the right contains additional information.	5/14	.02
2. Compared to the statement on the left, the statement on the right is more positive in it's tone and intent.	7/14	.01
3. The statement on the right is more concrete than the statement on the left.	0/14	.99
4. The statement on the right changes the meaning or intent of the statement on the left.	2/14	.99

ABSTRACT

HOW PROGRAM STAFF CAN PROVIDE FALSE EVALUATION DATA: THE PROJECT SEED EXPERIENCE

Keith McNeil and Charles Blanchard New Mexico State University

Evaluators usually assume that the data they receive are valid, or at least that the data reflect that being evaluated. In a recent evaluation in a large Southern school district evaluators had an opportunity to compare "raw interview notes" with "edited interview notes." The edited notes were judged by the evaluators to be so very different from the raw notes that the interview data was basically not included in the final report. The present study used raters who had no knowledge of the situation and hence had no vested interest in the outcome.

Each rater rated the original and edited pairs of comments on each of four scales. These scales were chosen because they were suspected to be the major ways that the edited responses differed from the original:

1. contains additional information.
2. more positive in tone and intent.
3. less concrete.
4. changes meaning or intent.

In order for evaluation data to be useful to decision makers, the data must be valid. Allowing those who are being evaluated to take control of the data and possibly modify the data may result in the data becoming more positive. If data is suspected of being manipulated, then the evaluator has the right, indeed the obligation, to not report those data.

C:\wp50\90mwsee