DOCUMENT RESUME

ED 383 567 SE 056 517

AUTHOR	Adey, Philip S.
TITLE	The Effects of a Staff Development Program: The Relationship between the Level of Use of Innovative Science Curriculum Activities and Student Achievement.
PUB DATE	95
NOTE	11p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; Elementary Education; Evaluation Methods; *Faculty Development; Foreign Countries; *Formal Operations; *Inservice Teacher Education; Intermediate Grades; Intervention; Program Evaluation; *Research Design; Science Education; Thinking Skills
IDENTIFIERS	England; Levels of Use of the Innovation; Process Product Research

ABSTRACT

This paper describes one aspect of a multi-factor investigation of the parameters which mediate in the effectiveness of an inservice staff development program. The paper also represents an attempt to recognize and to meet the problems associated with the evaluation of inservice programs for teachers in terms of student outcomes. The inservice course consisted of a two year program used to introduce methods of Cognitive Acceleration through Science Education (CASE) in 13 schools, between September 1991 and July 1993. CASE is a two year intervention program used in grades 6 and 7 whose objective is the promotion of formal operational thinking. Measures made in the study include assessments of the level of use of CASE by the teachers, and measures of cognitive gains by the students. A strong relationship was demonstrated between the level of use of the CASE intervention methods, as reported by teachers, and the cognitive gains made by their pupils. It is claimed that this finding represents both a substantial confirmation of the effectiveness of the CASE inservice program, and a demonstration of how a staff development program may successfully be evaluated using a process-product design. Contains 18 references. (LZ)

The Effects of a Staff Development Program: The Relationship Between the Level of Use of Innovative Science Curriculum Activities and Student Achievement

by
Philip S. Adey

The effects of a staff development program: the relationship between the level of use of innovative science curriculum activities and student achievement.

Philip Adey, King's College London Centre for Educational Studies

Process-product research revisited

At the 1994 NARST meeting Ellis, Enochs, & Mattheis, (1994) highlighted the need for quality evaluation of teacher development programs. Joyce & Showers, (1988) have argued strongly that if the purpose of a staff development program is to improve the quality of instruction, then ultimately the only validation of such a program must be evidence of improved learning by students of the teachers who have been exposed to the staff development program. Such a view is intuitively attractive, and yet attempts to evaluate staff development in terms of student achievement have run into many problems, which include those general to process-product research, and those particular to the evaluation of staff development.

Process - product research has become rather unfashionable. Richardson (1994) makes the case that much investigation by researchers in the past has been very instrumental in nature, treating teachers almost like objects to be manipulated, in a vain search for sets of teacher behaviours which can be relied upon to deliver good student learning. As a reaction to such ethically and psychologically dubious practices, the trend in classroom research has shifted towards ethnographic studies of classroom ecologies. Here I would like to look again at this question, and make the argument that while ethnographic studies have value for certain purposes, both socio-political and professional voices are quite reasonable in requiring some measure of outcome from investment in staff development, and that process-product research not only can yield useful information, but is the only approach which can in principle provide guidance to teachers and teacher educators on how professional practice might be changed to yield higher student achievement. Firstly, some of the specific criticisms of process-product research should be considered.

Doyle (1977), criticises studies in which specific teacher behaviours are correlated with student outcomes for the idiosyncratic way in which particular behaviours are chosen for study, and the unwarranted assumption of causality underlying the correlation. He compares the process-product paradigm unfavourably with the 'classroom ecology' paradigm:

3

"...the purpose of the ecological paradigm ... is to build and verify a coherent explanatory model of how classrooms work, a model that can be used to ask questions and interpret answers about teacher effectiveness"

It is clear that ethnographic studies of classrooms can - at a cost - provide far richer accounts of what happens in classrooms than can simply quantitative studies (see for example, Tobin, 1990). But whilst such studies provide rich descriptions, it is less clear how they can lead to prescriptions, that is, to advice to teachers or teacher educators about ways of improving their practice.

Fenstermacher (1979) also makes much of the problem of causality. He exemplifies the point with correlations found between, for example, the use of probing follow-up questions by the teacher and student achievement. He concludes reasonably that there is no way of telling from this correlation whether it is the nature of the questions that causes enhanced achievement, or whether higher achieving students provide feedback to teachers which encourages them to use higher level questioning techniques. Such criticism can be met by intervention studies, in which a teacher behaviour postulated as causally related to student achievement is specifically introduced, and changes in student outcomes observed. Fenstermacher's main criticism, however, is that process-product researchers necessarily, and unconsciously, make assumptions about what counts as "good" education. He claims that quantitative researchers are unaware that the products they strive for are no more than culturally determined norms. Bbut how important is such awareness? If teachers, students, parents, college admissions tutors, and employers all agree that high school grades or SAT scores, although crude, are the best measures available of achievement and aptitude then it seems to me that aiming for higher grades is a perfectly respectable aim for teachers and teacher educators. It follows that evaluation of inservice programs for teachers whose aims are the development of instruction must always, finally, look for evidence of increased student performance.

A further criticism of process-product research is the problem of interaction between particular teacher behaviours and particular learner personalities or learning styles, or context, which makes generalisation of results from individual studies difficult. In an elegant study, Gardner (1974) showed how the use made by different pupils of a given teacher behaviour was mediated by personality, such that the application of a simple process-product model could easily lead to erroneous conclusions. Where a particular teacher characteristic at first sight appeared unrelated to pupil performance, deeper analysis showed that it positively affected pupils of one personality type, and negatively affected pupils of a different personality type.

Brophy & Good (1986) in a thorough review of process-product research recognise all of these problems, and after eliminating studies which fail to meet their rather stringent criteria for acceptability, conclude

> "Despite the importance of the subject there has been remarkably little systematic research linking teacher behaviour to student achievement. A major reason for this is cost." (p.329)

They mean, of course, the cost of thorough and well designed studies. They also note that many studies are inconclusive because the student was used as the unit of study rather than class. In spite of the problems, however, Brophy and Good find that with more sophisticated observation methods and experimental designs, some reliable relationships began to be established between certain teacher attitudes and behaviours (such as warmth, business-like manner, enthusiasm, organisation, variety, clarity, structuring comments, probing follow-up questions, and focus on academic activities) and students' achievement. They conclude that process-product research is viable, but that it is difficult and requires careful attention to experimental design and interpretation to make its findings valid and usable.

Even if criticisms of process-product research can be met, two problems particular to staff development, which have received less attention in the literature, remain. The first is the dilution effect. An inservice staff development program can only be one of many influences on teachers, and a particular teacher can be only one of many influences on the students. The effect of one particular staff development program is likely to be so diluted in its effect on students as to be undetectable.

The second is the difficulty of isolating sources of failures of an inservice program to effect students. Much of what we do in inservice courses is based on unsupported assumptions about what constitutes effective teaching and learning. The measurability of outcomes associated with such assumed good practice presents a problem. If you are not sure whether or not teaching method X works, in any sense, then evaluation of an inservice program designed to introduce method X which shows no gain in pupil learning may either be because the inservice program was poorly delivered, or because method X does not work. There is no way of telling which.

Both of these problems can, in principle, be overcome: by making the staff development program sufficiently rich and dense so that its effect is substantial, and by evaluating the methods being advocated separately and establishing that, at least under optimum conditions, they can indeed lead to enhanced student achievement. Because such approaches tend to be expensive and time consuming, it is far easier to resort to 'evaluation' based simply on questionnaires asking teachers to make subjective judgements about the effects of the staff development program on their own performance and on their pupils' learning. It should not surprise us if such evaluation is met with scepticism by funders.

In this paper I attempt to recognise and to meet the problems associated with the evaluation of inservice programs for teachers in terms of student outcomes, and to show how attention to measures student outcomes which have wide acceptability, linked to teacher and curriculum inputs chosen on the basis of well-articulated psychological theories, which are specifically introduced to teachers can reveal causal relationships in which interaction effects are controlled. The establishment of such relationships can form the basis of specific advice to school principals and teacher educators about ways of improving student achievement through staff development. We should not allow the interesting ethnographic work done in the investigation of teacher thinking to distract us from the main business of evaluating staff development programs in terms of their measurable effects on students.

Context

This paper describes just one aspect of a multi-factor investigation of the parameters which mediate in the effectiveness of an inservice staff development program. The inservice course consisted of a two year program used to introduce methods of Cognitive Acceleration through Science Education (CASE) in 13 schools, between September 1991 and July 1993. CASE is a two year intervention program used in grades 6 and 7 whose objective is the promotion of formal operational thinking. The teacher processes built into the curriculum materials and which are the explicit subject matter of the inservice program include the generation of cognitive conflict in students, teacher and peer mediation to resolve this conflict leading to students' construction of reasoning patterns, and the encouragement of metacognitive reflection by students on their own conflict-resolution processes. The normal practice is for one CASE activity (Thinking Science, (Adey, Shayer, & Yates, 1989)) to be used every two weeks, instead of a regular science lesson. The CASE program has been shown to be consistently effective (Adey & Shayer, 1994; Shayer & Adey, 1993) in its aim to accelerate the development of formal operational thinking, and in turn to lead to long-term gains in students' academic achievement.

The inservice program was set up rather hurriedly in 1991 in response to a sudden demand from schools following the national publication of the achievement gains achievable with the CASE intervention and the min purpose was staff development rather than research as such. Nevertheless sufficient evaluative data became available to indicate the possibilities of this type of research.

The teachers of about 95 grade 6 and 7 classes were involved in the inservice program. Factors investigated in the whole study include measures of the teachers' sense of ownership of the innovation (SOO), the extent to which they communicated with one another in the school about the innovation (COM), the involvement of senior management in the implementation of the program (SMI), the Level of Use of CASE by individual teachers (LoU) and the effect size of student gains in cognitive development. At the last NARST meeting I reported (Adey, 1994) on some preliminary results obtained from this study, and showed that communication between teachers' within a school was directly related to their levels of use of the innovation. This paper reports on the effect of the inservice teacher program on students' development.

The staff development program consisted, over the two year period, of a total of six days of workshop and instruction activity held in a university department of education, and about 20 hours per school of coaching by the staff development tutors in the classrooms of the participating schools. This program answers both of the problems described above as common to the evaluation of inservice, since it was both extensive, and the CASE methods have been shown independently to lead to long term effects on students' levels of cognitive development and on their academic achievement. Questions of causality are answered, since the teaching method was specifically introduced and subsequent changes in student achievement measured. The class is used as the unit of analysis, and classes entered into data analysis are a random selection from a large sample very varied in terms of location, school type, and catchment area, thus generalisation from the results my be made with some confidence.

Measures

Measures made in the study which are relevant to the present paper are assessments of the level of use of CASE by the teachers, and measures of cognitive gains by the students.

The first of these variables is measured using the Loucks-Horsley Levels of Use (LoU) scale (Hall & Loucks, 1977) which relies on a structured interview linked to a rating scale for the responses. It yields a level on a scale from 0 (is not using and has no plans to use) to 6 (re-evaluation of the use including major modifications to increase impact). LoU interviews were conducted by two researchers with a sample of teachers drawn from each school on the basis of 'at least one-third, with a minimum of three from each school' with a selection made in each school to represent the full range of scores on the SOO measure. The two researchers cross-rated a sample of the taped interviews to ensure rating consistency.

The basic instruments used to obtain measures of cognitive gains are Science Reasoning Tasks (SRTs) (Shayer, Wylam, Küchemann, & Adey, 1978) ; see also Shayer, Adey, & Wylam (1981) for validation data. All SRTs are scored on a common scale in which early concrete operational thinking has a score of 3, and mature formal operations a score of 9. SRT II, Volume and Heaviness, was given to many of the classes involved in this study at the beginning of the intervention program and SRT III, Pendulum, was used at the end. Raw cognitive gains were obtained for each pupil by subtracting pre-test scores from post-test scores. For each class, the mean gain and standard deviation were obtained, and then these values compared with national norms for pupils of the same age and starting level (Shayer, Küchemann, & Wylam, 1976; Shayer & Wylam, 1978). The difference between mean gains over two years of CASE classes and national norms, divided by the standard deviation of the gains, gives the effect size of the enhanced gain over the normative gain. It is these effect sizes which are treated as the dependent variable in this study. In previous work (Adey & Shayer, 1993) we have established the relationship of cognitive gains made during the two years of the intervention program to subsequent increases in academic achievement assessed by national public examinations.

Results

Out of a total of 95 classes in 13 schools involved in the study, Level of Use data was obtained for 40 teachers, and complete and reliable pre-test and post-test data from 35 classes. Unfortunately the overlap between these sub-sets was not good, and for a total of only 18 classes from 7 schools was it possible to obtain both LoU data and cognitive gain effect sizes. Table 1 summarises this data. The Pearson product-moment correlation coefficient (which is robust and tolerant of quite wide deviation from the parametric, and also makes the best possible use of all of the data) between LoU and effect size is 0.611 ($p<.01$). The relationship is illustrated in figure 1.

Conclusion

In spite of the small size of the sub-sample for which complete data was available , and the limited range of values of both effect sizes and level of use available for analysis (in the whole sample effect sizes ranged from -1.12 to + 1.53, and levels of use from 0.5 to 6.0.), a strong relationship has been demonstrated between the level of use of the CASE intervention methods, as reported by teachers in a structured interview near the end of an extensive inservice program, and the cognitive gains made by their pupils. An inservice program has introduced teachers to the processes of CASE (cognitive conflict, metacognition, etc.) and these have been shown to be causally and quantitatively related to student products in terms of cognitive gains. It is claimed that this finding represents both a substantial confirmation of the effectiveness of the CASE inservice program, and a demonstration of how a staff development program may successfully be evaluated using a process-product design. There are thus both substantive and methodological lessons to be learned from this study.

References

Adey, P. S. (1994). The effect of school and departmental variables on the implementation of a science teaching innovation: paper presented to Annual meeting of the National Association for Research in Science Teaching; Anaheim, CA.

Adey, P. S., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. Cognition and Instruction, 11(1), 1 - 29.

Adey, P. S., & Shayer, M. (1994). Really Raising Standards: cognitive intervention and academic achievement. London: Routledge.

Adey, P. S., Shayer, M., & Yates, C. (1989). Thinking Science: The curriculum materials of the CASE project. London: Thomas Nelson and Sons.

Brophy, J., & Good, T. (1986). Teacher behaviour and student achievement. In M. Wittrock (Ed.), Handbook of Research on Teaching (pp. 328-375). New York : Macmillan.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. Shulman (Ed.), Review of Research in Education (pp. 163-198). Washington: AERA.

Ellis, J. D., Enochs, L. G., & Mattheis, F. E. (1994). Needed! Quality Evaluation of Teacher Enhancement Projects. paper presented to Annual meeting of the National Association for Research in Science Teaching; Anaheim, CA.

Fenstermacher, G. D. (1979). A philosophical consideration of recent research on teacher effectiveness. In L. Shulman (Ed.), Review of Research in Education (pp. 163-198). Washington: AERA.

Gardner, P. L. (1974). Research on teacher effects: critique of a traditional paradigm. British Journal of Educational Psychology, 44(2), 123 - 130.

Hall, G. E., & Loucks, S. F. (1977). A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 14(3), 238-237.

Joyce, B., & Showers, B. (1988). Student Achievement through Staff Development. New York: Longman.

Richardson, V. (1994). Conducting Research on Practice. Educational Researcher, 23(5), 5-10.

Shayer, M., & Adey, P. S. (1993). Accelerating the development of formal operational thinking in high school pupils, IV: Three years on after a two-year intervention. Journal of Research in Science Teaching, 30(4), 351-366.

Shayer, M., Adey, P. S., & Wylam, H. (1981). Group tests of cognitive development-ideals and a realisation. Journal of Research in Science Teaching 18 (2) 157-168.

Shayer, M., Küchemann, D. E., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. British Journal of Educational Psychology, 46, 164-173.

Shayer, M., & Wylam, H. (1978). The distribution of Piagetian stages of thinking in British middle and secondary school children. II - 14- to 16- year olds and sex differentials. British Journal of Educational Psychology(48), 62-70.

Shayer, M., Wylam, H., Küchemann, D. E., & Adey, P. S. (1978). Science Reasoning Tasks. Slough: National Foundation for Educational Research. (available now from Science Reasoning, Over, Cambs, UK.)

Tobin, K. K., J.B. Fraser, B.J. (eds.) (1990). Windows into Science Classrooms. Lewes: The Falmer Press.

Table 1 Level of Use (LoU) of CASE by individual teachers and the mean cognitive gain of their classes expressed in standard deviation units as effect sizes (eff size).

School	Teacher	LoU	eff size
2	201	4	0.40
6	601	5.5	1.37
8	802	5	1.12
8	804	3.5	0.22
9	902	5	0.83
9	904	3	0.70
9	907	3	0.94
11	1101	4	0.20
11	1103	4	0.53
11	1105	3	0.12
11	1105*	3	0.25
11	1114	4	1.31
12	1202	5.5	1.53
12	1204	5	0.90
13	1204*	5	0.56
13	1303	4	0.63
13	1305	5	0.97
13	1305*	5	0.61

* same teacher took two classes

Figure 1: LoU against effect size for 18 classes.