

# What Makes a Difference in Inservice Teacher Education? A Meta-Analysis of Research



*Staff developers need to take a second look at coaching and voluntary participation, among other things, when determining what really matters.*

RUTH K. WADE

Trying to keep abreast of the rapidly growing research on inservice education is a nearly impossible task.<sup>1</sup> Even if one can keep up with it, the results reported are often speculative, contradictory, and confusing. The research techniques, types of measurements, and the groups studied vary greatly among researchers. Drawing conclusions from such a heterogeneous conglomeration can lead to frustration and further confusion. Yet most research reviews and integrative works continue to be largely a pattern of reviewers' personal judgments, individual creativity, and preferred styles (Jackson, 1978).

Even though many more people are writing about staff development, few accounts present concrete evidence of its effects on teachers and students. Loucks and Melle (1982) conclude that most staff development reports are simply statements of participant satisfaction, which are then used to determine the success of a program. Baden (1982) suggests that inservice education should become a systematic effort to create behavior change in teachers and, eventually, in students. While participant satisfaction and local support are invaluable to inservice programs, there is a need to systematically determine their efficacy. Effectiveness should be measured not only at the level of the teacher-participant, but also at the level of the students with whom teachers interact.

*Ruth K. Wade is Staff Development Coordinator, Amherst-Pelham Regional Schools, Amherst, Massachusetts.*

*Author's Note: I am grateful to Professor Bill Wolf, University of Massachusetts, for his helpful critique of this article.*

What is needed is a systematic method for integrating findings across independent studies by converting them to a common base. Integrative analysis, or what Glass (1976) termed meta-analysis, provides the necessary perspective for this systematic integration.

With meta-analysis, all quantitative studies on a particular topic can be compared by making them part of a larger class or category: in this case, effects of inservice teacher training programs. The technique has limitations in that the researcher can only examine the data that commonly exist in a particular body of literature. The result, however, is precise quantitative estimates of the size of effect of various practices, based on a large number of studies.

Effect sizes are computed from summary statistics reported by authors. More specifically, effect size is the difference between the means of the treated and control groups divided by the standard deviation of the control group (Glass, 1977). By standardizing the scores, characterizing each as the size of the effect in relation to the standard deviation, the various sets of scores can be put on the same scale. After effect size calculations are made, statistical analyses can be used to identify the independent variables that account for the changes. A positive effect indicates that the experimental treatment was more effective than the control procedure, while a negative effect indicates that the control treatment was more effective. An effect size of .5 means that the treatment group showed one-half a standard deviation greater change than the control group.

In this study, I used meta-analysis to answer the following questions:

1. What are the effects of inservice training programs on teachers in the typical study?

2. Does inservice training have different effects for different intended instructional outcomes? For example, does the effect size vary if the goal of the training is increasing participant learning, changing participant behavior, eliciting positive participant reactions to the training, or influencing students of participants?

3. Finally, do the effects of training vary as a function of duration of training, training group characteristics, location and scheduling, sponsorship, incentives for participation, training group structure, and instructional techniques?

## Method

I began by making an exhaustive search of the literature to obtain studies that met the topical criteria as well as provided the necessary data from which to calculate effect size. I reviewed over 300 journal articles, dissertations, and Educational Resources Information Center (ERIC) documents published or presented between 1968 and 1983; of these, 91 were selected for inclusion in this meta-analysis. To be included, a study needed to meet the following criteria: (1) the study was quantitative rather than qualitative; (2) the data necessary for calculating effect size were presented; (3) the data related to one of the major questions in this study; and (4) subjects of the study were public school teachers or their students in grades K-12.

I initially read many of these selected studies to determine the variables that were frequently represented in this body of literature and to ferret out the most commonly used independent variables. I came up with a list of 28 variables, which I grouped into eight categories in order to describe the main features of the studies. These variables include: (1) effect levels or goals of the training, (2) duration, (3) training group characteristics, (4) location and scheduling, (5) sponsorship, (6) participant incentives, (7) structure, and (8) instructional techniques. It is important to note that while I examined each variable independently, I also looked at variables 2 through 8 in terms of variable 1 to determine whether the goal of the training influenced the effect size of each variable.

I treated each study as an individual unit of analysis and considered multiple dependent variables or multiple comparison groups as separate data sets. The 91 studies yielded 715 data sets.



## Results

1. *Effect Levels.* For all studies, I separated effect levels into four categories that defined the level at which the evaluation was directed:

• *Reaction* assessed how the participants felt about the inservice training.

• *Learning* measured (usually through pretests and post-tests) the amount of learning that was achieved.

• *Behavior* measured whether participants changed their behavior as a result of a staff development intervention.

• *Results* determined whether there was an impact in the classroom, usually on students, as a result of teacher training.

The findings indicate that inservice teacher education programs reported in the literature are moderately effective. Inservice treatment of any kind, on the average, resulted in .52 of a standard deviation greater change than control groups. Such an outcome should provide reassurance to staff developers and other educators who wonder whether the time, money, and effort invested make a difference.

When the data are grouped by the level at which the evaluation was directed, the findings are conclusive. Attempts to increase participants' *learning* through inservice teacher training are highly effective (.90 mean effect size); attempts to change participants' *behavior* and to elicit positive *reactions* to the training are moderately effective (.60 and .42 mean effect size, respectively); while attempts to demonstrate *results* by looking at the students of participants are only mildly effective (.37 mean effect size).

2. *Duration.* The studies in this meta-analysis indicated that there was no significant effect of length of treatment. This includes studies with training ranging from a few hours to those

lasting more than 30 hours. Similarly, there was no significant difference in the effect sizes of programs lasting six months or less versus those lasting more than six months.

3. *Training Group Characteristics.* Training groups involving both elementary and secondary teachers achieved higher effect sizes than groups enrolling only elementary or

only secondary level educators. My findings corroborate the often-reported fact that greater effects are evident after training a group of elementary teachers versus a group of secondary teachers, but working with both groups together may suggest some new training possibilities.

Contrary to popular opinion, whether a participant voluntarily chooses to



**"...the evidence is beginning to point to the fact that coaching, as an instructional technique, does not have the potential to alter teacher behavior...."**



attend inservice training or is required to attend does not make a significant difference in training effect size. It is reassuring to note that over 85 percent of the 609 data sets in this category reported voluntary participation; in fact, they had 23 standard deviations greater effectiveness than mandatory participation studies, but the difference was not statistically significant.

The size of the training group (1-20, 21-40, 41-60, greater than 60) did not have a significant effect on training results, nor did composition of the group in terms of whether participants were a faculty unit or a group of unrelated individuals.

4. *Location and Scheduling*. No location or scheduling variables produced a statistically significant impact on effect size. Variables included on-site versus off-site training and scheduling during or outside of school hours.

5. *Sponsorship*. Training programs initiated, developed, or funded by the state or federal government or a university were significantly more effective than those initiated within the school, either by teachers, administrators, or supervisors. This finding does not support the popular belief held by many staff developers that teacher-initiated training programs are more

effective. Perhaps education professionals who work outside the schools have more time and financial resources available to develop, test, and present training programs for teachers. In any case, these results do not support the limited federal role in school improvement suggested by Berman and McLaughlin (1978), but do support the School Improvement Study (Loucks, 1983), which found high success rates in federal- as well as state-funded programs.

6. *Participant Incentives*. When a participant was selected to take part in training, either by being designated as a representative of a particular group



or through a competitive selection process, the effect size was significantly greater than for all other incentives studied. This effect may take place because the "best" people are chosen to participate, or perhaps these individuals work harder when they have been "chosen" to participate.

The incentive of college credit, followed by released time, produced moderately positive effect sizes. Pay, certificate renewal, and no incentive resulted in small positive effect sizes that were well below the mean effect size for all incentives studied.

7. *Structure*. Independent study produced the highest effect size of the structures examined. An explanation for this large effect may be that the independent study structure appeals to the most highly motivated people. There do not appear to be important differences in the effect sizes among workshops, courses, mini-courses, or institutes, all of which produced effect

**"Contrary to popular opinion, whether a participant voluntarily chooses to attend inservice training or is required to attend does not make a significant difference. . . ."**

sizes near the mean and are moderately effective.

8. *Instructional Techniques*. The results of analyzing data by the types of instruction that yield the highest effect sizes indicate that four types of instruction are significantly more effective than others. The most effective instructional methods are observation of actual classroom practices, micro teaching, video/audio feedback, and practice.

Instructional methods associated with significantly lower effect sizes are discussion, lecture, games/simulations, and guided field trips.

In spite of many speculative claims that coaching greatly enhances instructional effectiveness, I could find no evidence for this in the 225 cases using coaching. Coaching, modeling, mutual assistance, printed material, production of instructional materials, programmed study, and film were all moderately effective but did not produce significantly higher effect sizes than the mean of all instructional methods examined.

The next logical question to examine is whether some combination of instructional methods produces significantly higher effect sizes than the mean of all instructional methods examined. I grouped together many different combinations of types of instruction to see if some particularly effective combination existed. These combinations included: (1) lecture, modeling, practice, coaching; (2) lecture, discussion, modeling, practice, coaching; (3) lecture, discussion, modeling, video/audio feedback, coaching; and many others. The result was that any combination resulted in an effect size somewhere near the mean attained for all instructional methods grouped together. Thus it appears that there is no "magical" combination of methods for effective instruction.

Practical rather than theoretical instruction, with the instructor taking almost exclusive responsibility for the design and teaching of the class, results in significantly higher effect sizes. In classes where participants are encouraged to teach each other through classroom presentations, group work, and discussion sessions, a lower effect size results.

Although instructors represented many job categories (such as teacher, support staff, administrator, consultant, college personnel), only self-instruction produced a large positive effect size. This may say something about the effectiveness of self-instruction as a technique, but it more likely indicates that any individual who is motivated enough to complete a program of self-instruction is likely to achieve successful results. Support staff and college personnel as instructors are moderately effective, while teachers and state department of education representatives produced only small positive effects. Administrators and consultants did not have enough cases for making significant conclusions.

### Conclusions

Meta-analysis provides an objective technique for research synthesis. Its emphasis on quantification ensures against projecting personal bias onto a vast field of educational research. Most of my findings are consistent with earlier meta-analyses of inservice training by Joslin (1980) and Lawrence and Harrison (1980). However, I examined several new categories that had not been previously examined through meta-analysis. My findings contradict some commonly held assumptions regarding effective training practices.

There is no "magic formula" for

effective inservice programs. Figure 1 provides a summary of practices associated with statistically significant, above-average effectiveness. Staff developers who wish to plan programs for maximum effectiveness should contemplate the following suggestions, which are based on the outcomes of this meta-analysis:

1. Plan programs in which elementary and secondary teachers can participate in training together whenever appropriate.
2. Encourage teachers to become involved in state-, federal-, or university-initiated programs.
3. Offer incentives for participation, such as enhanced status or college credit, whenever possible.
4. Encourage independent study and self-instruction as alternatives to the traditional workshop format.
5. Suggest that instructors set clear goals and take major responsibility for the design and teaching of the class rather than encouraging participants to assume these roles.

6. Use instructional techniques such as observation, micro teaching, video/audio feedback, and practice as alternatives to lecture, discussion, games/simulations, and guided field trips.

Throughout the staff development literature, coaching has been cited as an effective technique for achieving "transfer of training." As Griffin (1982) noted, "this inferential model is a potentially powerful one for staff developers if it proves to be accurate." However, the evidence is beginning to point to the fact that coaching, as an instructional technique, does not have the potential to alter teacher behavior as proposed by Joyce and Showers (1981). Not only has this study found that coaching to achieve transfer of training was only moderately effective, other evidence concludes that coaching as an instructional technique may not always be effective. Sparks (1983) found that workshops plus trainer-provided coaching were not superior to workshops alone or to workshops plus peer observation.

**Figure 1. Inservice Practices Associated With Above-Average Learning Effectiveness.**

Practice	Effect Size
Attempts to increase participants' learning	.90
Federal, state, and university initiated programs	.69
Training groups with elementary and secondary teachers	.67
Participants who were selectively chosen to participate	.76
Self-instruction	.92
Independent study	.98
Strong leadership roles by instructors	.66
Instructional techniques:	
Observation	.81
Micro teaching	.78
Video/audio feedback	.64
Practice	.55

Levinson (1962) suggested that the "psychology of coaching" is an important, yet often neglected consideration, which results in the failure of coaching in the field of management. Levinson cited four reasons that coaching might fall short of its potential as a development technique:

- The coach and the trainee rarely have the "psychological time" to develop the kind of relationship based on mutual respect that is necessary for effective coaching.

- Because there is usually not much tolerance for giving people time to grow, the coaching relationship is often impaired by pressure from superiors to get information from coaches that might be used against the trainee.

- Coaches often do not know how to foster independence; therefore, quick solutions are sometimes proposed that do not fit the complexity of the problem.

- The coaching situation is in danger of being blocked by the universal

feeling of rivalry and its accompanying fears.

These potential impediments may be why coaching was only moderately effective in the 225 cases in this meta-analysis. Under some specific circumstances, coaching may be an effective technique, but it does not seem to provide a panacea for staff development programs.

Although these suggestions should guide staff developers, they are only a part of what is needed to provide effective inservice education. Many factors are involved in determining the effectiveness of any given inservice activity. Context issues—such as understanding the school climate; principal support; and adequate resources, including time and an understanding of the needs—should not be underestimated. Nor should process issues such as governance and teacher investment be overlooked.

Nevertheless, the odds for successful staff development can be increased

if staff developers consider the factors that have been clearly demonstrated to be related to effectiveness. □

Inservice education and staff development have been used interchangeably to mean any training activity designed to increase the competencies needed by teachers in the performance of their assigned responsibilities.

#### References

Baden, D. J. "A User's Guide to the Evaluation of Staff Development: Assessing the Impact of Staff Development Programs." Syracuse, N.Y.: National Council of States on Inservice Education, 1982.

Berman, P., and McLaughlin, M. *Federal Programs Supporting Educational Change, Vol. VIII: Implementing and Sustaining Innovations*. Santa Monica, Calif.: The Rand Corporation, 1978.

Glass, G. V. "Integrating Findings: The Meta-Analysis of Research." In *Review of Research in Education*, Edited by L. S. Shulman. Itasca, Ill.: Peacock Publishing Company, 1977.

Glass, G. V. "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher* 5 (1976): 3-8.

Griffin, G. A. "Staff Development." Paper prepared for the NIE Teaching Synthesis Conference, Arlie House, Virginia, 1982.

Jackson, G. B. "Methods for Reviewing and Integrating Research in the Social Sciences." Final report to the National Science Foundation for Grant No. DIS 76-20309. Washington, D.C.: Social Research Group, George Washington University, 1978.

Joslin, P. A. "Inservice Teacher Education: A Meta-Analysis of the Research." Doctoral dissertation, University of Minnesota, 1980.

Joyce, B. R., and Showers, B. "Transfer of Training: The Contribution of Coaching." *Journal of Education* 163 (1981): 163-172.

Lawrence, G., and Harrison, D. "Policy Implications of the Research on the Professional Development of Education Personnel: An Analysis of Fifty-Nine Studies." In *1981 Report on Educational Personnel Development*. Washington, D.C.: Feistritzer Publications, 1980.

Levinson, H. "A Psychologist Looks at Executive Development." *Harvard Business Review* 40 (1962): 69-75.

Loucks, S. F. "At Last: Some Good News from a Study of School Improvement." *Educational Leadership* 41 (1983): 4-5.

Loucks, S. F., and Melle, M. "Evaluation of Staff Development: How Do You Know It Took?" *The Journal of Staff Development* 3 (1982): 102-117.

Sparks, G. M. "Inservice Education Training Activities, Teacher Attitude, and Behavior Change." Doctoral dissertation, Stanford University, 1983.

## Highlights of Research on Inservice Teacher Education

A meta-analysis of 91 well-documented studies shows that:

- Inservice training that includes both elementary and secondary teachers is often more effective than inservice for either group separately.

- Inservice is most successful when participants are given special recognition for their involvement, are selected on a competitive basis, or are designated to participate.

- Regardless of who conducts inservice sessions (trainers come under many different job classifications), teachers are more likely to benefit when they learn on their own. Similarly, of all the different types of training structures, independent study is the most effective.

- There is no magical combination of methods for successful inservice. Nevertheless, inservice programs that use observation, micro teaching, audio and visual feedback, and practice—either individually or in some combination—are more effective than programs that do not use these methods.

- There is no evidence that "coaching" greatly enhances instructional effectiveness. At best, it is moderately effective.

- Inservice is less successful when participants are regarded as major contributors. Programs are more effective when the leader assumes the role of "giver of information" and the participants are "receivers of information."

Copyright © 1984 by the Association for Supervision and Curriculum Development. All rights reserved.