



The Link Between Staff Development and Student Achievement

“There is no question that staff development can raise student achievement when it addresses the academic content that teachers teach, their teaching repertoire, and the amount of practice they provide students in particular areas.” (Bruce Joyce, director of Booksend Institute, in Sparks, 1998)

Demonstrating the link between staff development and student achievement challenges most evaluators. Although this connection may seem obvious, the proof that staff development leads to increased student achievement eludes evaluators. The link between staff development and student achievement is both intuitively strong and methodologically challenging.

Starting the Conversation

Results-Based Staff Development for the Middle Grades was launched to answer the question: Which staff development programs improve student learning? The National Advisory Panel discovered almost immediately that this work would generate more questions than answers. The myriads of questions were daunting at times; however, the painstaking process of considering all the relevant questions has made possible this collection of examples of staff development programs that provide evidence of impact on student learning. The National Advisory Panel hopes that this work will lead to further dialogue and learning that will benefit not only those who seek to answer the question for various constituents, but also students who ultimately benefit from teachers' learning.

What Works in the Middle: Results-Based Staff Development will help others who are trying to discover which staff development programs demonstrate an impact on student achievement. This chapter explores the challenges of evaluating staff development and summarizes the evaluation methods used by the programs included in the guide. The chapter also addresses the systemic nature of staff development and how it affects evaluation processes. In addition, the chapter discusses the difficulties of attempting to prove that staff development increases student achievement. Finally, the chapter discusses how the programs included in this guide have demonstrated that staff development influences student achievement.

Systemic Nature of Staff Development

To incorporate only staff development in an effort to improve student achievement is to tinker around the edges. Staff development is certainly necessary to increase stu-



dent achievement. However, staff development cannot be successful unless the system in which it occurs supports high levels of learning for both staff and students. When staff development is present — along with other factors that support quality staff development and student achievement — students' achievement increases.

Staff development is much like the respiratory system in the body. As one of the body systems, it is essential to the body's basic operation. But, to be fully functioning and healthy, the body needs all its systems working together. Removal or dysfunction of any system leaves the body in poor health and at risk. The same is true for school improvement efforts focused on increasing student achievement. To be successful, school improvement requires multiple systems to work together to succeed. These systems include staff development, compensation, teacher evaluation, student assessment, and many others. Eliminating any one system increases the risk that school improvement effort will be unsuccessful.

In addition, simply knowing that teachers participated in staff development and that student achievement increased does not prove that staff development was responsible for the increase. Multiple factors such as higher standards, improved curriculum frameworks, and new types of assessment are also associated with increased student learning. No one factor alone leads to increased student achievement. Because they are integrated simultaneously within a school system, none of these factors, including staff development, can be measured in isolation. It is nearly impossible in the complex social system of schools to determine if a particular factor was exclusively responsible for increased student achievement. Therefore, staff development leaders and decision-makers need to acknowledge the relationship of many factors rather than to attempt to show that staff development is a single cause of increased student achievement.

The evaluation of the programs in this guide is correlational, not causal. The programs in the guide demonstrate that a positive relationship exists between staff development and student achievement. However, a cause and effect relationship has not been verified. Staff development was present in all of the cases where student achievement was realized and is certainly one "systemic" element related to the documented increase in student achievement in each of the programs.

Evidence Not Proof

Rigorous experimental research to provide proof that staff development causes increases in student achievement is not possible in the complex social environment of schools. Too many intervening variables occur simultaneously, especially in schools engaged in systemic reform. If proof is not possible, Guskey (1998) suggests that evaluators of staff development collect evidence about the impact of staff development. Joyce (Sparks, 1998) suggests that we stop trying to select that elusive, "perfect" form for academic evaluation of staff development efforts. It is quite pos-

sible that new forms of evidence and new approaches to evaluation will need to be applied to demonstrate the link between staff development and student achievement. Instead, at this point, staff development leaders, researchers, and practitioners need to put on the table for discussion the issues about and examples of evaluations that demonstrate the impact of staff development on student achievement.

What is evidence of impact? This question appears simple but is laden with embedded values and beliefs. Prior to answering this question, evaluators need to understand that different audiences may want different answers to this question. For example, teachers may want to know how much effort a student expends on a particular academic task. Principals may be interested in knowing if students are coming to school and attending classes. Policy and decision-makers may want to know what the return on the investment is for expenditures in staff development. And, some audiences may not be interested in isolating staff development as the single factor that improves student achievement. Instead, they may be satisfied by simply knowing that when a school provides additional resources for reading, increases the instructional time for reading, and provides staff development designed to help teachers more effectively use the increased instructional time, student reading achievement increases. Responding appropriately to these various needs requires different forms of evidence and more flexible research designs.

Knowing what a school's, or district's, diverse audiences want to know about the relationship between staff development and student achievement will guide evaluators. Evaluators then need to select an appropriate research design and collect appropriate evidence of student achievement. Without baseline information about what information is needed, and for which audiences, evaluators will have a difficult time planning assessments.

What constitutes appropriate evidence of student achievement? The National Advisory Panel posed its own questions about what constitutes good measures of student achievement. For example, are standardized achievement tests with a standard error often exceeding five months powerful enough to measure increases in student learning? Or, what forms of assessment will measure increases in student achievement that result from changes in teacher content knowledge and instructional practice (e.g. greater use of inquiry or using writing in mathematics or science)? What evidence best demonstrates increases in student achievement? Must there be a standardized test, or will performances or authentic products, which meet prescribed standards, be sufficient to document student achievement?

The primary criterion for any project to be considered for inclusion in *What Works in the Middle: Results-Based Staff Development* was evidence of student achievement — what students know and are able to do. For the purpose of this study, indicators of

student achievement include measures such as norm-referenced tests, student portfolios, performance tasks, state assessments, local criterion-referenced tests, and increased enrollment and success in advanced-level courses. A full discussion of the measures of student achievement appears in Chapter 3, “The Selection Process.” While these indicators are related strictly to students’ academic success, evaluators might also evaluate whether their program goals require them to consider other indicators such as increased attendance, participation in class, satisfaction with school, or sense of self-confidence.

Evaluation Designs

Research designs to measure the impact of staff development on student learning are typically quasi-experimental or qualitative rather than experimental. Experimental research design allows the researcher to control for extraneous factors — those differences that exist in the subjects and environment that may influence changes in student achievement. It also requires random assignment of subjects to control and treatment groups. When staff development is implemented school-wide or district-wide and students are in intact classes, randomization is not feasible. The approach most similar to strict randomization is to assign teachers and classrooms to either experimental or control groups or to identify equivalent groups through statistical equalization.

The most common form of evaluation used in the 26 programs included in this guide is quasi-experimental. Quasi-experimental research is a form of experimental research done when the subjects are not randomly assigned to treatment and control groups. Qualitative research was used in one program. In qualitative research, researchers describe, interpret, and explain events in the real world. Of the 26 programs included in this guide, quasi-experimental and qualitative research designs were used in all but five cases. The exceptions randomly assigned classrooms and teachers, not students, to either a treatment or control group.

Some researchers who used quasi-experimental research accommodated for potential differences between control-treatment groups prior to the treatment. They conducted statistical measures of equivalency to demonstrate that both the control and treatment groups were similar. This process provides some compensation for the lack of random assignment to control and treatment groups.

Several research designs were used to demonstrate the link between staff development and student achievement. Table 1 presents the various evaluation designs used to demonstrate the link between staff development and student achievement in the 26 programs included in this guide. Along with a brief description of each design are the specific programs that used each evaluation design. If multiple measures of impact were conducted, some programs are listed more than once. The data sources or measures of student achievement are listed for each design. Table 1 also comments upon the strengths and limitations of each design.

Table 1: Evaluation Designs

Design	Data Sources/ Measures	Strengths	Limitations
Experimental			
<p>1. Pre-post test with randomly assigned control/comparison and treatment groups (random assignment of teacher and/or classes)</p> <p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • CRISS • Project LEGAL • Project Success Enrichment • Six-Trait + 1 • Student Team Literature 	<ul style="list-style-type: none"> • norm-referenced tests • program-developed tests • state assessments 	<ul style="list-style-type: none"> • measures growth • permits a calculation of significance • increases the generalizability of results • reduces the chance that the change is the result of other factors • accounts for differences in the groups before treatment • increases the ability to isolate the effects of staff development 	<ul style="list-style-type: none"> • requires advanced planning • may not be possible to randomly assign groups in real-life contexts • results may be affected by pre-test (testing and sensitizing effect)
Quasi-Experimental			
<p>2. Post-test only with non-equivalent/matched control/comparison and treatment group</p> <p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • ECRI • FAST • HALP • Math Renaissance • National Writing Project • Rice University Student Mathematics Project • SPAN • SWRP • We the People ... Project Citizen 	<ul style="list-style-type: none"> • norm-referenced tests • program-developed tests • state assessments • performance assessments with established scoring guides 	<ul style="list-style-type: none"> • measures changes in achievement • permits a calculation of significance • eliminates testing effects (practice and sensitizing) 	<ul style="list-style-type: none"> • does not account for difference in the groups prior to the treatment • requires advanced planning • may be difficult to select or identify a control group • does not account for other factors that may have contributed to the growth
<p>3. Post-test only with equivalent/matched control/comparison and treatment groups</p> <p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • Junior Great Books • We the People ... The Citizen & the Constitution 	<ul style="list-style-type: none"> • norm-referenced tests • program-developed tests • state assessments 	<ul style="list-style-type: none"> • measures changes in achievement • increases the generalizability of results • reduces the chance that the change is the result of other factors • reduces testing effects 	<ul style="list-style-type: none"> • may be difficult to identify a control/comparison group • requires advanced planning • does not account for differences in the groups prior to treatment

Quasi-Experimental

<p>4. Pre-post test with no control/comparison group</p> <p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • EarthStorm • Fernwood Project • PUMP • RUSMP 	<ul style="list-style-type: none"> • program-developed tests • state assessments • performance assessments with established scoring guides 	<ul style="list-style-type: none"> • measures changes in achievement • permits a calculation of significance 	<ul style="list-style-type: none"> • requires advanced planning • does not account for extraneous factors • does not permit generalizability to other programs • results may be affected by the pre-test (practice and sensitizing effect)
<p>5. Pre-post test with nonequivalent/ matched treatment and control/comparison groups</p> <p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • Iowa Chautauqua • National WritingProject • Profile Approach to Writing • Reading Power in the Content Areas 	<ul style="list-style-type: none"> • norm-referenced tests • program-developed tests • state assessments • performance assessments with established scoring guides 	<ul style="list-style-type: none"> • measures growth • permits a calculation of significance • increases ability to isolate the effects of staff development 	<ul style="list-style-type: none"> • control/comparison and treatment groups may differ prior to treatment • results may be affected by the pre-test (practice and sensitizing effect)
<p>6. Pre-post test with equivalent/matched control/comparison and treatment groups</p> <p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • ELOB • Junior Great Books • Powerful Connections • RUSMP <p>7. Case study</p>	<ul style="list-style-type: none"> • norm-referenced tests • state assessments 	<ul style="list-style-type: none"> • measures growth • permits a calculation of significance • increases the generalizability of results • reduces the chance that the change is the result of other factors • accounts for differences in the groups before treatment 	<ul style="list-style-type: none"> • changes may be the result of the pre-test (practice and sensitizing effect) • may be difficult to identify a control group • requires advanced planning

Qualitative

<p><u>Programs using this design:</u></p> <ul style="list-style-type: none"> • Introducing Math Teachers to Inquiry 	<ul style="list-style-type: none"> • performance assessments with established scoring guides 	<ul style="list-style-type: none"> • describes changes that occur as a result of the intervention 	<ul style="list-style-type: none"> • does not account for other factors that may have contributed to the changes • does not permit generalizability to other programs
--	---	--	---

Body of Persuasive Evidence

The search for persuasive evidence to demonstrate the link between staff development and student achievement was one goal of the Results-Based Staff Development for the Middle Grades initiative. The search resulted in identifying 26 staff development programs with evidence of increased student achievement. Individually these efforts may not be persuasive; however, as a collection of studies across a wide span of subject areas, in many diverse settings, and with different measures of student achievement, they provide convincing evidence that staff development is strongly related to student achievement.

Even though the relationship between staff development and student achievement is logically and intuitively sound, identifying a body of evidence to support that a strong relationship exists is not easy. Additional evidence to support this body of research is important. Evaluators, staff development leaders, and program coordinators must join forces to monitor, gather additional evidence, and communicate the results of their work to extend the evidence presented in this guide.

Limitations of This Work

The studies included in this guide have a number of methodological flaws and, in some part, are evidence of a single year's results rather than multi-year, longitudinal studies. What they do represent are significant attempts to answer the question: Does content-based, results-based staff development for middle-grades teachers increase student achievement?

While *What Works in the Middle: Results-Based Staff Development* does not provide conclusive proof to support the link between staff development and student achievement, it provides evidence that there is a strong link between them. Further, it suggests that additional study of appropriate ways to demonstrate this relationship is necessary. These staff development programs help construct an answer to the question: Does staff development make a difference? What they do not help us know is how much difference it makes. Nor does this work answer questions about what aspects of the staff development program contribute most to teacher and student learning. There are strong patterns or similarities among these programs described in Chapter 5, "Common Characteristics of Programs in the Guide." Yet more research is needed to determine if these similarities are responsible for the success of the programs included in this guide.

To build additional support for the hypothesis that teacher learning increases student learning, both practitioners and researchers must expand the body of evidence drawn with other evaluations from disparate situations, identify the best ways to document the increased student achievement, and determine if it is possible to demonstrate how much staff development impacts student learning.



WALK ... DON'T RUN

Lifeguards at neighborhood pools, where eager kids gather to enjoy the cool, refreshing waters on a hot summer day, spend almost all day saying: "Walk; don't run" or "Slow Down!". Children, in their enthusiasm to get into the water, often disobey the rules posted on fences and in locker rooms. Too often, educators, in their enthusiasm to initiate innovations to improve student performance need the same cautions and reminders:

WALK; DON'T RUN. SLOW DOWN!

References

Guskey, T. (1998, Fall). The age of our accountability. *Journal of Staff Development*, 19, 36-44.

Guskey, T. (in press). *Evaluating Staff Development*. Thousand Oaks, CA: Corwin Press.

Sparks, D. (1998, Fall). Making assessment part of teacher learning: An interview with Bruce Joyce. *Journal of Staff Development*, 19, 33-35.