

**Making Do With Less: Interpreting the
Evidence from Recent Federal
Evaluations of Dropout-Prevention
Programs**

**Mark Dynarski
Mathematica Policy Research, Inc.
December 2000**

**Prepared for a conference on “Dropouts:
Implications and Findings” at Harvard
University, January 13, 2001.**

**I wish to thank Mary Moore and Paul
Decker for helpful comments and
suggestions.**

MAKING DO WITH LESS: INTERPRETING THE EVIDENCE FROM RECENT FEDERAL EVALUATIONS OF DROPOUT-PREVENTION PROGRAMS

Beginning in the late 1980s, the U.S. Department of Education conducted three large evaluations of the effectiveness of programs to reduce dropping out. The programs and the evaluations were supported by funds from the Carl Perkins Vocational Education Act and two phases of the School Dropout Demonstration Assistance Program (SDDAP), one operating from 1989 to 1991, the other from 1991 to 1996. Together, the three evaluations studied more than 100 dropout prevention programs and rigorous evaluation designs were used for 30 of these programs.

Findings from the three evaluations show that most programs did not reduce dropping out by statistically significant amounts, but some programs did improve some outcomes. Three programs (funded in the second phase of the SDDAP) that prepared students who had already dropped out to get the General Education Development certificate improved GED completion rates. An alternative high school on a community college campus reduced dropout rates. And several alternative middle schools reduced dropout rates.

The three evaluations were broad-ranging studies and two of the three relied on random assignment techniques to measure program effects reliably. Considering the extent and rigor of these evaluations, do their findings comprise a menu of program approaches that a policymaker or education program developer could use to select an effective dropout-prevention program for their school or district?

In this paper, I argue that we do not yet have a menu of program options for helping students at risk of dropping out. The evaluation findings are useful as guides to further program development and testing, but they fall short of providing a scientific basis for implementing programs in new schools or districts based on the models. Recognizing the urgency of the issue, however, I suggest an alternative approach to identify approaches for helping at-risk students that program developers can use while efforts to develop a stronger scientific basis for programs continue.

The approach I suggest puts a premium on the ability of a program developer to readily see or infer the “logic model” inherent in an education idea or approach being considered. The logic model is the statement of the pathways by which a program will achieve its objectives. According to the approach I suggest, programs are more desirable when it is clear how they can be expected to affect teaching or learning, or keep students in school. Doubts or confusion about how a program will achieve its objectives should be viewed as a downside to the program. I note the elements of dropout prevention programs to which their effectiveness may be traced and suggest that implementing these elements—rather than “a program”—may be a useful strategy to reduce dropping out.

A Summary of Key Evaluation Findings

The largest and longest of the three evaluations focused on programs funded by the second phase of the School Dropout Demonstration Assistance Program. The evaluation studied 20

programs in depth, collecting data on almost 10,000 students for up to three years. Experimental designs were used for 16 of the 20 programs (which were termed “targeted” because students meeting particular criteria were targeted for program services). The other four programs were school-wide reform efforts that were evaluated using comparison-student designs.

Random assignment is a powerful method. It compares what happens to program participants (technically, treatment-group members) to what happens to students who are *statistically equivalent* to program participants (technically, control-group members). These students were eligible for the programs but were denied entry as part of the evaluation. Experiences of equivalent students are a proxy for what would have happened to program participants if they had not been able to enter the program.¹ Among the 16 programs, 8 programs served middle-school students and 8 served high school students.

Summary results for the 16 programs are presented in Tables 1 and 2.² Among the eight middle school dropout prevention programs, half provided low-intensity supplemental services such as tutoring or occasional classes to promote self-esteem or leadership.³ Four middle school programs in the evaluation took a more intensive approach to serving at-risk students. Two of these programs--the *Griffin-Spaulding Middle School Academy* near Atlanta, Georgia and the *Accelerated Academics Academy* in Flint, Michigan--were alternative middle schools with facilities that were physically separate from the regular district middle schools. The other two programs--*Project COMET* in Miami, Florida, and *Project ACCEL* in Newark, New Jersey--were located within regular schools but separated students from other students within the school for much of the day. These four programs typically taught students in smaller classrooms than regular middle school students and provided more intensive counseling services. Three of the four programs primarily served students who were overage for their grade level, and these programs attempted to accelerate students’ academic progress to allow them to “catch up” with their age peers.

Supplemental programs had almost no impacts on student outcomes. None of the programs affected the dropout rate, and average student grades, test scores, and attendance were similar among treatment and control group students (Table 1).⁴ The alternative middle school programs in the evaluation were more successful in keeping kids in school and accelerating their academic progress. Compared with control group students, treatment group students admitted to these programs were half as likely to drop out and completed an average of half a grade more of school (Table 1). On the other hand, alternative middle schools did not seem to help students learn more in school. Alternative middle schools in the evaluation had no impacts on grades or test scores, and they had impacts on attendance in the wrong direction (treatment group students were absent more often than control group students). Although students were promoted at a faster rate than students in regular middle schools, student learning did not seem to improve in these programs.

TABLE 1
IMPACTS OF MIDDLE SCHOOL DROPOUT PREVENTION PROGRAMS

	Average Treatment Group Mean	Average Control Group Mean	Number of Sites	Number of Sites with Significant Impacts
Supplemental Programs				
Dropout Rate (Percentage)				
End of Year 2	7.8	7.0	4	0
End of Year 3	11.5	15.0	4	0
Days Absent				
During Year 2	10.5	10.0	4	0
During Year 3	14.3	14.3	4	0
Math Grade				
Year 2	69.5	68.3	4	1(+)
Year 3	67.5	67.0	4	0
Reading Score (percentile)				
Year 2	36.0	35.5	2	0
Year 3	37.0	34.0	1	0
Alternative Middle School Programs				
Dropout Rate				
Year 2	4.7	9.3	3	1(-)
Year 3	9.0	18.0	2	1(-)
Highest Grade Completed				
Year 2	7.9	7.4	3	3(+)
Year 3	8.6	8.1	2	2(+)
Days Absent				
During Year 2	18.3	15.3	4	3(+)
During Year 3	18.0	17.0	2	0
Math Grade				
Year 2	65.0	66.3	3	0
Year 3	62.0	64.0	2	0
Reading Score (Percentile)				
Year 2	16.3	16.7	3	0
Year 3	28.0	31.0	1	0

SOURCE: Dynarski et al. (1998)

^aPlus and minus signs indicate whether impacts were positive or negative.

TABLE 2
IMPACTS OF HIGH SCHOOL DROPOUT-PREVENTION PROGRAMS

	Treatment Group Mean	Control Group Mean	Number of Sites	Number of Sites with Significant Impacts
Alternative High School Programs				
Dropout Rate				
End of year 2	35	30	5	1(+)
End of year 3	39	40	3	0
Completion Rate				
HS diploma	21	15	4	0
GED	13	19	4	1(-)
Either	33	34	4	0
GED Programs				
Dropout Rate				
End of year 2	56	58	3	0
End of year 3	57	60	3	0
Completion Rate				
HS diploma	9	3	3	0
GED	30	20	3	0
Either	39	24	3	1(+)

SOURCE: Dynarski et al. (1998).

NOTE: For alternative high schools, completion rates refer to the second follow-up year for two programs and the third follow-up year for two programs. For GED programs, completion rates refer to the third follow-up year.

^aPlus and minus signs indicate whether impacts were positive or negative.

Impacts in Atlanta and Flint

	ATLANTA		FLINT	
	Treatment Group	Control Group	Treatment Group	Control Group
Dropout Rate (Percent)	6	14	2	9
Highest Grade Completed	8.6*	7.9	85*	7.8
Math Grade	59	63	67	66
Reading Score (Percentile)	--	-	12	12

NOTE: All outcomes measured at the end of the second follow-up year, except for highest grade completed, which is measured at the end of the third follow-up year in Flint.

* Significantly different from the control group at the ten percent level, two-tailed test.

: Dynarski et al. (1998).

The effects of alternative middle schools were concentrated primarily in the Atlanta and Flint programs (see box). Evidence from Atlanta and Flint suggests that something positive happened for their students. On the sobering side, however, is the lack of effects on attendance and academic performance.

The high school programs were all intensive compared to the middle school programs. Five of the high school dropout-prevention programs in the evaluation offered high school diplomas, with four being alternative high schools and one being a school within a school.⁵ None of the five programs significantly lowered dropout rates (Table 2). However, alternative high schools seemed to influence whether students earned a diploma or a GED. In Four of the five alternative high school programs, more students earned high school diplomas and fewer earned GED certificates compared to control group students. The differences were not statistically significant in any of the four sites, but the pattern is consistent across sites. Control group students were less likely to earn a high school degree and more likely to earn a GED.

A closer look at Seattle’s *Middle College High School* provides insight about how alternative high schools can affect high school completion. Middle College High School had higher high school completion rates and lower GED completion rates (see box) for students whose characteristics suggested that they were least likely to drop out (termed “low risk” students in the box, though most were at some risk of dropping out). The school also reduced dropping out for high-risk students.

Impacts of Seattle’s Middle College High School

Seattle’s Middle College High School is an alternative high school on a community college campus. The program served dropouts or students on the verge of dropping out of regular high schools and screened students to ensure that they were motivated to succeed.

	Low-Risk Students		High-Risk Students	
	Treatment Group	Control Group	Treatment Group	Control Group
Dropout Rate	33	33	27*	42
Completion Rate	53	56	59	58
HS diploma	33	24	27	25
GED	20	32	32	33
In High School	13	11	13	0

NOTE: Outcomes are measured at the end of the third follow-up year. Percentages may not add to 100 because of rounding.

* Significantly different from the control group at the ten percent level, two-tailed test.

SOURCE: Dynarski et al. (1998)

One key feature of Middle College High School is that it had staff and current students interview prospective students to ensure that they were motivated adequately for the challenge of completing high school. The positive impacts of the school suggest that alternative high schools possibly can be successful when they serve students who want to succeed. Of course, some caution needs to be exercised in linking program impacts to any one program feature.

Three other programs offered GED certificates, with each being structured as a small alternative high school. Two programs in the evaluation--the Queens, New York *Flowers with Care Program* and the St. Louis, Missouri, *Metropolitan Youth Academy*--were designed to help students prepare for the GED, and a third program--the *Student Training and Re-entry Program* in Tulsa, Oklahoma--was a transition program for high school dropouts to help them determine and achieve an appropriate educational goal, which usually turned out to be a GED certificate. Table 2 shows that participants in the three GED programs were more likely to earn their GED certificates than control group students and even somewhat more likely to complete their diplomas than control group students (this result arises because students who start in GED programs can leave the program and go to other programs or back to high school). The total effect is that GED programs improved the overall high school completion rate from 24 percent to 39 percent, a relative increase of over 60 percent.

Among the three GED programs, the Metropolitan Youth Academy in St. Louis had the largest impacts (see box), with 39 percent of treatment group students earning a GED certificate or a high school diploma within three years, compared to 22 percent of control group students. This is a substantial effect, and it is especially notable since the academy served students who were more at risk than any other program in the evaluation.

Impacts of the St. Louis Metropolitan Youth Academy

St. Louis’s Metropolitan Youth Academy is a GED program for highly at-risk students. Nearly all of the students served were dropouts and had, on average, the most risk factors of any program in the evaluation. The program was more successful at helping students earn GEDs than other programs in the St. Louis area.

	St. Louis	
	Treatment Group	Control Group
Dropout Rate	60	66
Completion Rate	39	22
High school diploma	11	3
GED	28	19
Attending HS or GED program	2	11

NOTE: All outcomes measured at the end of the third follow-up year.

SOURCE: Dynarski et al. (1998)

Learning From Evaluations

Knowing that some programs have beneficial effects is a good start. From a scientific standpoint, the logical and careful next step would be to replicate an effective program in a variety of circumstances and possibly with a variety of different “tweaks.”

Two reasons to replicate a program are the contextual nature of program effects and the difficulty of implementing a program exactly to specification. The contextual nature of the effects arises because measured effects of the program depend on the experiences of the control or comparison group. A program impact is a relative concept, a difference in outcomes between two groups. The weakness of evaluation findings based on only a few sites is that the same impacts may not arise when a program is implemented in a different site with a different context for the control or comparison group (for example, the control group may have more or less services available). The value of testing the model in a range of settings is precisely so that the control or comparison group contexts can vary and the impacts can be measured against the varied contexts.

The second reason to replicate is to test the robustness of the model to changes in its services or approaches. A program may be successful because a constellation of intangible elements comes together in its favor, but whether the constellation can be replicated elsewhere is a key issue. The program may have had a particularly effective leader or staff, or outstanding support from district administrators. It may be possible to write down generally how a program should be implemented, but it is more challenging to follow instructions to “have an effective leader” or “ensure that district administrators support the program.” Replication is useful to for testing whether the unseen elements of a program are the crucial ingredient to its success. If they are, replicating the program’s success is unlikely.

It is rarely possible to implement a program exactly as a model or the program’s designers would specify (perhaps adaptation is a better term than replication). Programs serving young people typically consist of interrelated services and activities whose compositions depends on infrastructure, skills, and resources in the local area. A school-within-a-school program, for example, consists of a physical space, a targeting strategy for who should be served, and possibly smaller class sizes, a different curriculum, training for teachers, social services or counseling, and a modified governance or administrative structure. A school district wanting to implement a school-within-a-school may have a board with diverse views on serving at-risk students, a collective bargaining agreement that creates procedures governing how the program can be staffed, and physical or regulatory constraints on the particular space to be used for the program. With local circumstances possibly requiring compromises to the model, policymakers face the daunting task of identifying how to implement various elements, without clear guidance as to the relative important of the elements to the model’s success. It would not be surprising if programs often were implemented that were flawed copies or were otherwise less effective than their models.

Referring to the impact findings above, a strict interpretation of the findings is that the results for alternative middle schools are promising for program contexts similar to what the control groups experienced in Flint and Atlanta. To assess the ultimate promise of the programs, however, alternative middle schools based on the Flint and Atlanta models would need to be replicated in other districts and possibly with some variations in services or students served. An middle school model that was effective when implemented in a range of school districts, and probably with deviations from its initial model, would be a promising approach for helping at-risk middle school students. Evidence from replications sets a high standard and programs meeting the standard merit special attention.

By its nature, replication involves multiple units—schools, districts, or perhaps states—implementing a program. Considering how decentralized education is, however, higher-level government agencies, foundations, or partnerships need to assume responsibility for conducting replication research. And at the federal government level, which arguably has the most resources to devote to research, evaluations of education programs are increasingly common as a tool for understanding and developing better programs and for accountability.

However, many federal evaluations currently are much different from studies of particular models or replications of models. Typically, federally-funded education programs are not highly prescriptive about what states or districts need to do to operate the program, and, in response, local programs can attain a wide variety of shapes and sizes, even when being funded from the

same federal source. In this context, evaluations of federal programs tend to focus on whether the *federal* program is effective rather than on whether *local* programs are effective.

For example, the federal evaluation of the Title I program (Puma et al. 1993) was based on nationally representative samples of students in about 100 school districts and about 400 schools, and results are not reported at the district or school level. The national evaluation of the Upward Bound program (Myers and Schirm 1999) is based on a random sample of local programs, which means the findings can be generalized to the program as a whole, but the sample sizes for any one program are small (averaging about 50 students) and not designed to support inferences about which local program approaches are most effective. Moreover, because both the Title I and Upward Bound programs had operated for decades before these particular evaluations had begun, neither study could use a strategy of implementing models identified to be of special interest to the field, though it was likely that some schools and students in their samples were involved with innovative models.⁶ Partly to offset this feature, the Title I evaluation had a “special strategies” component that looked at exemplary models of school reform and the Upward Bound evaluation used qualitative research techniques to look more closely at “high-impact” programs.

The federal emphasis on the overall program has another implication. Federal evaluations rarely test various models relative to each other. The value of testing models relative to each other is that the tests yield evidence about winners and losers. However, testing models against other models would mean exercising far more control of the types of programs that are funded, and possibly even controlling the allocation of programs (such as by random assignment) to school districts or schools that carry them out. This more scientific funding strategy—in which programs are funded consistent with a strategy for replicating models—is consistent with the view expressed by Nobel Laureate James Heckman, who, with Jeffrey Smith, wrote that

The end result of a research program based on experiments is just a list of programs that “work” or “don’t work,” but no understanding of why they succeed or fail. The long-run value of cumulative knowledge is high, but is neglected by advocates of “short-run” evaluations conducted outside of coherent social-science frameworks. (Heckman and Smith 1995)

Note that the view expressed by Heckman and Smith applies to experiments per se, but if a program were being evaluated using a nonexperimental evaluation design, the same point applies. Their view is more a criticism of the role program evaluations have come to play rather than of the designs used for evaluation.

Addressing the Heckman-Smith criticism would be difficult in the current context. Scientific funding strategies and program granting strategies are quite different. Discretionary program grant strategies define criteria that local programs must meet. Reviewers then rate grant applicants based on the criteria, resulting in a list of applicants who have the highest ratings and who, subject to agency discretion to ensure geographic or other types of balance, receive the grants. Only by happenstance would funded applicants fit within a scientific strategy that would support testing particular models. Formula-funded grant programs, which generally direct money to states based on some count of individuals or students in a states, offer less discretion and therefore are less likely to meet scientific criteria.

Ensuring that applicants fit within a scientific strategy would mean identifying scientific criteria in advance and picking applicants to fit within these (and presumably other) criteria. The recent Interagency Education Research Initiative comes close to these criteria and is a promising approach for aggressively studying particular education topics (new IERI grants focus on foundational skills and teaching of complex math and science ideas in middle and high school).⁷ However, other new programs are created more to fulfill policy goals than scientific goals. Recent federal initiatives—such as Gear Up, programs supported under the Reading Excellence Act, 21st Century Community Learning Centers, and, more recently, the small learning communities initiative, have awarded grants in competitions that do not have scientific criteria. As such, any evaluation necessarily is coming in after the fact. The evaluations must try to learn what they can within the constraints of an existing set of programs that may or may not have given the needed balance to enable the evaluation to study particular hypotheses or models.

If replication research is not likely to emerge from the federal government, it may emerge from foundations. Some replication efforts are supported by foundations or nonprofit organizations, such as the Career Academies study being conducted by MDRC (Kemple et al. 2000) and the New American Schools initiative. Generally, however, examples are limited.¹

Making Do With Less

In the end, policymakers and program operators have to make decisions about programs and funding with less information, perhaps much less information, than the scientific ideal would suggest. Rather than a well-tested model of an alternative middle school, for example, the policymaker is presented with evidence of one or two alternative middle schools or GED programs that were found to be effective. Is this a basis for implementing these programs?

Faced with this situation, which I submit is common, policymakers have several options. They can do nothing, but the problem they want to address then is unabated. They can look for corroborating evidence from other sources. However, the newer the idea, the less likely it will be that other research can corroborate it. Or they can implement the program, and take the risk that it will work in their instance (or perhaps even work better).

It is this third option that I want to focus on here. Borrowing a principle from investment planning, I suggest that a program developer or funder considering a new program (a risky investment) should strive to understand how a program's structure or approach will yield the expected results. To give this idea more substance, I suggest trying to complete the following sentence for a program or approach being considered: "This program will help students learn or teachers teach because _____." Of course, other objectives could be used in the sentence, but certainly these two are a useful starting point. Because the question is simple, and for lack of a better term, I will call programs "simple" if the question can be answered for them. Programs that are not simple I will call "complex."

Using the test, I categorized a sampling of current ideas in education policy and research and arrived at the following table.

Simple	Complex
<ul style="list-style-type: none"> ■ small classes ■ small schools ■ extended learning time ■ teacher training ■ charter schools 	<ul style="list-style-type: none"> ■ whole school reform ■ school to work ■ school-community partnerships ■ technology

Why is class size reduction a “simple” program? Using the test sentence, my response is “class-size reduction will help students learn and teachers teach by increasing the time teachers can spend with each student, by increasing the degree of control teachers have over student behavior in the classroom, and by enabling teachers to use instructional strategies that might be impossible with larger classes.” If test scores or other measures of academic performance were to increase following a class size reduction, it is likely that some combination of the above factors is responsible for the increase.

Whole school reform is categorized as a “complex program.” Perhaps I am doing an injustice to whole school reform models, but my attempt to answer the sentence was unconvincing. “Whole school reform will help students learn and teachers teach because teachers will be able to use new teaching strategies (though training and professional development may be needed to ensure teachers can use the strategies), possible changes in school governance may lead to a smoother functioning school (though the link between governance and teaching and learning is vague), possible increases in parent involvement may signal students that parents and school staff share common values and have high expectations, (though increasing parent involvement is a challenge for any school)...” and there is more to consider. In the end, attempting to answer the question made whole-school reform seem like a black box. With so many aspects of a school being affected by the reform, it would be difficult to ever know what was the source of a positive effect.

One could say at this point that why the reform works is not important. If it works, it works. As I have argued, however, knowing that a reform has worked somewhere else is less useful than knowing whether it will work in a school or district where one wants to implement it. The fact that reforms cannot be implemented the same in other settings is a reason to focus on “simple” programs. For example, reducing class size is the same concept everywhere, though its magnitude could vary. It may work for somewhat different reasons from place to place (or not), but the concept itself can be clearly understood. This is not to say the concept is easy to implement, of course.

The table is not intended to convince readers that particular programs are simple or complex. Rather, I recommend the table as a tool for readers to create their own categorizations. It is instructive then to compare ideas on the left and the right and identify patterns. For my table, the so-called “simple” ideas on the left seem to have a more immediate relationship with teaching and instruction. Ideas on the right have classroom elements also (for example, technology

programs that put multimedia computers into classrooms affect teaching and learning, though how much is still not well understood) but include components that occur elsewhere in the school. School-community partnerships can be structured with no classroom component at all.

The table categorizes charter schools as simple ideas based on the ability to answer the test question. Charter schools may improve teaching and learning because the freedom to develop a school that suits the staff, parents, and students in exchange for accountability may result in more innovative teaching and more energetic learning. That the hypothesis underlying charter schools is simple does not mean it is simple to set up a charter school, or that the issue is not controversial or even divisive. There are many types of charter schools, an array of problems in sorting out start-up financing and staffing, and different forms of governance of charter schools, all of which contribute to implementation challenges.

Note that I am not asserting that more evidence exists for simple ideas compared to complex ideas. Indeed, nearly all the ideas in the table would benefit from further study and research, as the charter-school example suggests. A vigorous effort currently is underway to research whether charter schools can fulfill their promise and the research ultimately will yield more evidence whether charter schools are effective. However, charter schools can be implemented and what is being tested by a charter school is clear.⁸

I also am not asserting that whole school reform or other complex ideas are ineffective. Many of the reforms have shown substantial effects when implemented in some schools and districts. To date, however, a question remains about whether the demonstrated success of reforms is due to the model itself or to the nature of the school and staff undertaking the model. The distinction is that if the school and the staff are the source of the effects, it may not matter which reform is implemented, as long as one is. Random assignment of schools into treatment and control groups would help address the question, a strategy that is currently being pursued in a study of the Success For All program.

Implications for Dropout Prevention

Policymakers interested in reducing the number of dropouts no doubt would like to hear something they could do for which there is some evidence it will be effective. The pattern of evidence from the impact assessment points to programs that were effective. However, I am not suggesting that policymakers should implement versions of alternative middle schools based on the Flint and Atlanta programs or GED programs based on the Metropolitan Youth Academy in St. Louis. These programs might be good starts for replication efforts but one or two examples of effective programs is too narrow a base to stand on.

Rather, we can use the test question to identify features of the programs that contributed to their effectiveness. For the Flint and Atlanta schools, my test sentence is: “The alternative middle school programs in Flint and Atlanta helped keep students in school by creating small schools, with smaller class sizes, that helped teachers know students better and provide more help, and a focus on helping students address personal and family issues through counseling and access to social services.” Programs oriented to get students GED certificates had the largest impacts on high school completion (albeit through completing the GED). For example, the St. Louis program raised GED completion rates from 22 percent to 39 percent. For GED programs,

my test sentence comes as: “The GED programs in St. Louis, Queens, and Tulsa helped students complete the GED by providing individualized instruction in a small setting with access to counseling and social services.”

What is evident in comparing the two sentences is that effective programs operated in small settings and paid attention to students needs inside and outside the classroom. Generally the programs were smaller than regular schools and students had more access to adults who could help them with issues and problems. Moreover, site visitors frequently interviewed program staff who described how they wanted to help and went out of their way to provide help. The programs recognized that students often had family or personal problems that hindered their ability to attend or succeed in school, and tried through counseling or other means to help students deal with the problems. And the programs recognized that students needed a measure of academic challenge, that even students with undistinguished academic records could respond to teachers pushing them to learn, especially when learning somehow was connected to their personal experiences. Indeed, reviews of other dropout prevention programs have noted that successful programs generally have this characteristic of *personalization* (Fashola and Slavin, 1998).

However, it is also true that other programs in the evaluation tried to create small settings and pay attention to student needs but were not effective. No impacts were observed for an after school program to build literacy skills, a peer mentoring program, a homework assistance and tutoring program, and several schools within schools. These results illustrate why replication is valuable. If the results in the effective programs were due to other factors that were more difficult to observe than small settings and personal attention, a replication would likely show less positive findings. In the meanwhile, the findings here suggest that a high degree of personalization—a strategy of focusing intensively on why students are having difficulty and actively working to address the sources of the difficulties—is worth considering.

REFERENCES

- Dynarski, M., P. Gleason, A. Rangarajan, and R. Wood. *Impacts of Dropout Prevention Programs*. Princeton, NJ: Mathematica Policy Research, 1998.
- Dynarski, M., and P. Gleason. *How Can We Help? What We Have Learned from Evaluations of Federal Dropout Prevention Programs*. Princeton, NJ: Mathematica Policy Research, 1998.
- Fashola, O. and R. Slavin. "Effective Dropout Prevention and College Attendance Programs For Students Placed At Risk." *Journal of Education for Students Placed at Risk*, vol. 3, No. 2, 1998, 159-183.
- Hayward, B., and G. Talmadge. *Strategies for Keeping Kids in School: Evaluation of Dropout Prevention and Reentry Projects in Vocational Education*. Raleigh, NC: Research Triangle Institute, June 1995.
- J. Kemple and J. Snipes. *Career Academies: Impacts on Student Engagement and Performance in High School*. Executive Summary. New York: Manpower Demonstration Research Corporation, 2000.
- Myers, D, and A. Schirm. *The Impacts of Upward Bound: Final Report from Phase I of the National Evaluation*. Washington, DC: Mathematica Policy Research, Inc., 1999.
- Puma, M., C. Jones, D. Rock, and R. Fernandez. *Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity*. Interim Report. Washington, DC: Abt Associates Inc., 1993.
- Rossi, R. *Evaluation of Projects Funded Under the School Dropout Demonstration Assistance Program*. Palo Alto, CA: American Institutes of Research, March 1994.

ENDNOTES

¹Two features of the evaluation's design affect interpretation of the results. First, students in the treatment group did not always enter or stay long in the dropout prevention program being studied. By using the full treatment group and control group, instead of just program participants, the evaluation measured the impact of *access to* dropout prevention programs, which may underestimate the impact of *participation in* dropout prevention programs. Data collected for the evaluation were not sufficient to distinguish the two different impact estimates. Second, students in the control groups were able to receive other dropout-prevention services available to them. Thus, the measure of program impacts reveals how the program affects students relative to other programs in the area.

²Dynarski and Gleason (1998) present the highlights of the results and findings from the evaluation. Dynarski et. al (1998) present the complete findings from the evaluation.

³These programs included the Albuquerque, New Mexico, *Middle School Leadership Program*, the Chula Vista, California, *Twelve Together Program*, the Long Beach, California, *Up With Literacy Program*, and the Rockford, Illinois, *Early Identification and Intervention Program*.

⁴Table 1 shows average student outcome levels among treatment and control group students across the supplemental middle school and alternative middle school programs in the evaluation. Since data were not always available from every site, the table also shows the number of sites on which the treatment and control averages for a particular outcome are based. The table also shows the number of sites for which impacts were statistically significant at the 10 percent level. In reporting the results, we focused on cases in which there were statistically significant effects in a large number of sites.

⁵The four sites operating alternative high schools were Boston (*JFY High School and University High*), Las Vegas (*Horizon High Schools*), Miami (*Corporate Academy*), and Seattle (*Middle College High School*). The school within a school approach was used in Chicago (*Wells Academy*).

⁶A notable exception is the U.S. Department of Labor's replication study of the Quantum Opportunities Program, which began in 1996. Findings from the study are expected in Summer 2001.

⁷The announced goal of the IERI is "to improve preK-12 student learning and achievement in reading, mathematics, and science by supporting rigorous, interdisciplinary research on large-scale implementations of promising educational practices and technologies in complex and varied learning environments." The program announcement indicates that "only those projects that meet high standards of methodological rigor, are of sufficient scale, integrate technology, and are conducted by interdisciplinary research teams will be funded." The program announcement can be found at <http://www.ehr.nsf.gov/ehr/rec/ieri/default.htm>.

⁸Policy often does not wait for research to generate findings. Even as the Clinton administration proposed spending a billion dollars a year to reduce class sizes (formally, the

initiative was to hire more teachers), researchers were debating whether reducing class sizes would improve education performance (Federal Reserve of NYC conference). An interesting aspect of the class-size reduction initiative was that the administration initiated no primary research to study whether class sizes improved achievement using experimental or quasi-experimental techniques. Instead, it relied on syntheses of existing findings that gave a prominent role to a class-size experiment conducted in Tennessee in the late eighties. It is at least questionable whether the findings in Tennessee generalize to the nation as a whole, and the class size reduction initiative proposed by the administration involved reductions in class size that were much smaller than the reductions studied by the Tennessee experiment.