

CHOOSING SUCCESS

BY SAM STRINGFIELD

THERE IS no shortage of programs that promise to turn around low-performing schools, but how can you tell which ones will live up to their claims? The key is to approach the choice of a school reform program as an important and complicated consumer decision. Schools or school systems will have a better chance of making a good choice if they ask these three questions about any program they are considering:

■ First, are the goals and objectives of the program in line with the goals and objectives of your school or school district? No matter how sound a program is, or how successful elsewhere, it will be a waste of everyone's time and effort if it is not designed to get your school where you want it to go.

■ Second, how strong is the research supporting the program's claims of success? Answering this question, which hinges on a number of technical issues, is likely to be the most daunting of the three. It is also essential. As Al Shanker often observed, none of us would use a medicine that had not been found safe and effective in rigorous research. Why should we expect anything less of the programs we hope will reform our schools?

■ Third, given the money you have to spend and the people you have to work with, is a given plan practical for you? A program might come with the best possible pedigree in terms of research design and prior success, but if the financial or human costs are beyond the resources of your school or district, the design is of no real value to you.

Making any kind of major change is hard work. A program that is no good—or not right for your school—involves just as much work as one that would fit your

Sam Stringfield is principal research scientist at the Johns Hopkins University Center for the Social Organization of Schools, Baltimore, Maryland.

school like a glove and help you raise your students' achievement. What follows is a kind of road map to help schools or districts attempting reform to be intelligent consumers. I'll pay particular attention to the question of the research base for a design, as that is the most technical and the least a matter of common sense.

1. A Program's Goals and Objectives

The first job in evaluating the appropriateness of a program is to see how closely it matches your goals and objectives. To do this, you may have to cut through the marketing verbiage of brochures and videos and the rather abstract statements that are the currency of the school reform movement to uncover the program's specific goals and objectives. They should be readily translatable into working hypotheses that you can test. Statements like "All children can read by the age of nine" or "Today's students must be prepared for 21st century jobs" sound impressive and are certainly valuable, but they don't tell you enough about what a program proposes to accomplish and how. In Maryland (Tennessee, Illinois, Florida, etc.), where every school will be judged by its mean scores on the MSPAP (TCAP, IGAP, etc.), the questions you need answered are, "What effects can hard-working practicing professionals expect this design to have on our MSPAP (TCAP, etc.) scores?" and "Over what period of time?"

Or, if your MSPAP scores are already above the district or state average and you're worried that your students aren't learning enough about other cultures (or aren't attending regularly enough or whatever the goal of your reform happens to be), the question again is exactly how this design proposes to help you. Does it offer more help than would be offered by an alternative design? What *measurable* changes will you see in students' achievement or knowledge or attendance? Are they in line with your expectations?

2. A Program's Research Base

The next and far more difficult question is, "How strong is a program's research base?" What kind of proof is there that a program will live up to its claims? In some cases, the research base will be very sturdy. The program will have been put into practice in a number of schools over several years or even longer, and careful data about results will have been kept and analyzed. I call a program that meets these stringent criteria an A-level or A-list program, and I describe how you can recognize an A-level program in more detail in the following section.

In 1998, there are still very few programs that meet these criteria, and it would be impractical to limit people's choices to A-level programs. So, in two subsequent sections, I provide descriptions of programs with increasingly less sturdy research bases—they are B-level and C-level programs.

A solid research base, while extremely important, is not the only criterion for choosing a program, so you would not necessarily be making a mistake in choosing a C-level program over an A-level one. The issue of a good match in terms of goals, already discussed, and the issue of practical viability, which I'll discuss in the final section, are also important in making a choice. Nevertheless, a program's research base provides essential information about how likely the program is to perform as advertised. All other things being equal, a school should take an A-level design much more seriously than a C-level.

What about programs that do not even meet C-level standards? There are a large number of such programs being successfully marketed in the U.S. today; they may even be in the majority of school reform programs. These designs should be avoided, however, until better research on their effects is available.

How to recognize an A-level program. A strong practical research base would include the following:

- A number of studies of the program's effectiveness, preferably carried out by independent experts.

- Detailed information about the kinds of students the program is designed to serve. For example, did the students attend urban, Title I, schoolwide project schools? Schools in middle-class suburbs or in small rural school districts? All of the above?

- Several studies that include carefully matched control groups. These might chart the progress of the program schools compared with progress in other local schools that are demographically similar, comparing, for example, reading scores (or math scores or student attendance or whatever) in program and control schools.

- Studies in which indicators of success, determined in advance, are plausibly related to what the program has promised to deliver. For example, a program designed to raise scores might use gains on a widely used reading test as a standard of success.

- Several studies that are at least two, and preferably three or more, years in duration. (As an example, see *Special Strategies*, Stringfield et al., 1997.)

- In a majority of the studies, proof that program schools produced educationally significant student gains in the target areas (e.g., student achievement, attendance, graduation rates, rates of disciplinary referrals, or whatever else was promised). The gains would be represented as either moderate-to-large effect sizes (E.S. = .4 or greater) or experimental-control means that certainly appear educationally significant, combined with tests of the statistical significance ($p < .01$) of differences between experimental and control groups. Effect sizes in this range almost always translate into achievement gains greater than 10 percentiles. Ideally, the studies would provide both effect sizes and tests of statistical significance. If your school improvement team does not include a person who has studied statistics, you should recruit technical assistance from your district's central administration, a federally funded consolidated center, or a nearby university.

- A number of carefully conducted case studies. This is a plus rather than a necessity. The case studies would include both the strong points of the program and the problems that became apparent during its implementation, as well as the reactions of the various groups involved in the implementation described.

- An even-handed article written by a third party, reviewing multiple studies of the effects of the reform. Such an article is not essential to the credibility of an A-list program, but if one exists, it should be taken quite seriously. For example, Fashola and Slavin (1997) review studies of several promising programs and, based on what the studies show, divide the programs between those that are well documented and researched (for our purposes, an A list) and designs that are "promising" but have not yet been rigorously researched. These latter designs would belong on our B list or C list. Stringfield et al. (1997, chapters 10-12), and Ellis and Fouts (1997) provide shorter but somewhat overlapping sets of designs and programs. These could be similarly valuable to persons seeking research reviews.

The designers of the program must be willing to provide a full list of participating schools, their names, addresses, and phone numbers. (Ask about schools that are no longer participating, as well.) Using a set of questions that members of the selection team have agreed upon, call five participating schools at random and ask about their experiences (cost: under \$20 and under two hours). You should get a strongly positive set of reviews about the design's feasibility and its effects on students. Pay particular attention to the reviews from schools that are most like your own. Of course, the comments from former participants, if any, should also be added to the mix. For decisions that are this practical, a seeing-is-believing test is important. Visits to at least two program schools should give you a lively sense of the challenges involved in putting the program into effect and an equal sense of its potential benefits.

In short, an A-level program has undergone multiple, rigorous evaluations, and the evaluation studies provide clear information on implementation and clear outcome measures, and they show significant gains as a result of participation. With an A-level program, you are unlikely to find many troubling examples of failure.

Examine your options aggressively. If you choose a reform program, implement your choice relentlessly.

There is currently no reform program that is a perfect A. However, the Success for All/Roots and Wings (Slavin et al., 1996) design comes closer than most others to meeting this definition, as do a few of the designs described in the review provided by Fashola and Slavin (1997) and Stringfield et al. (1997).

B-level programs: less of the same. Since in 1998, the list of A-level programs is very short, a school may also need to consider (with caution) promising programs from the B list. B-level programs are still in the top 10 percent to 20 percent of all school reform programs. A program that looks promising may not be on the A list because it has not yet been the subject of rigorous, well-controlled, longitudinal studies of the kind described in connection with A-level programs. This could be a matter of choice (not a good sign), but since such studies are expensive, it could be that the developer has not yet found the money to finance them. Also, a new design will, by definition, lack the longitudinal data necessary for these kinds of studies.

The fact remains that B-level programs do not have the kind of evidence of success that the A-level programs have. Given the weaker research base, credible B-level programs should offer the following:

- Two or more studies reporting student outcome data, each lasting at least one year, and preferably at least two. Or the research base might include a half-dozen or more rigorously conducted case studies carried out over several years.

- Positive results in cases where the program has been used in schools or school districts like your own. This is especially important when the data are limited.

- Alternatively, well-documented studies (like those described under the A list) showing that the design tends to produce positive but relatively small effects on desired student outcomes (e.g., an effect size greater than .2, or at least a 5 percentile gain).

- Data from process-only evaluations (e.g., "the students and teachers enjoyed the program" or "a majority of the parents expressed great satisfaction" or, better, "measures of student engagement indicated a significant rise in stu-

dents' involvement"). These data can be considered, with the caveat that they don't take the place of clear student-outcome data: The fourth-graders who say they love a program could still be reading at a second-grade level.

- As at the A level, the designers of a program should be willing to provide a full list of participating schools, their addresses, and phone numbers. Since the proof that these programs can produce what they promise is not as strong as it is with A-level designs, the calls and follow-up visits are even more important. A random calling of seven of these schools (cost: under \$30 and under three hours) should produce a positive set of reviews of the design, its practicability, and its effects on students. Again, pay particular attention to the reviews from schools like your own and visit the two nearest schools that are serving communities similar to yours.

You should be very cautious about any reform that has been in existence for five or more years and has not met B-level research specifications. The absence of adequate data is just as damning as data pointing consistently in the wrong direction. Not every reformer wants to become a psychometrician; but any group asking school people to spend tax dollars, and the time and work of teachers and others, must understand that its word that a program will work is not enough.

C-level programs: One could be right for you. What if neither the A list nor the B list offers a program that is just right for your school? You may find a new program that has only been tried in a few schools but is interesting and looks like a good fit with your school and your goals. Or perhaps one or more influential people in your school are passionate believers in a particular C-list program and seem willing to work relentlessly to make it successful. Under these circumstances, you could choose such a relatively untried design and have a chance of success, but only if the program has the following bare-minimum research support:

- Strong links to convincing research in areas related to the program. (For example, the program might involve research-proven types of cooperative learning.) A vague claim based on research that does not have much to do with the operation of American public schools (like one asserting that the program is related to recent "brain research") should be regarded with skepticism. So should a simple assertion of the developer's expertise, however impressive. (He or she may be an expert, but that does not prove that this particular program will work.)

- Two to five published case studies demonstrating the positive effects of the design in schools like yours.

- An open acknowledgment by the design team of any research or case studies that found mixed or negative student outcomes and a willingness to discuss why you might expect different results in your school.

- A list of all schools currently using the program. Since the formal research is not strong, it is all the more important to make calls and follow-up visits to program schools. A random calling of ten of these schools (cost: under \$40 and under one day) should produce a set of reviews that

are nearly all positive as to the design of the program, its implementation requirements, and, above all, its effects on students. Pay particular attention to reviews from schools like your own.

■ A frank discussion of the characteristics of schools that have discontinued participation in the design. Ask the designers for the names of these schools and discuss their reasons for leaving the program with the schools as well as with the developers. While this step is important with any school reform design, it is essential with a C-list program.

A C-level design should be chosen only if it provides an excellent match to the needs of a particular school. No matter what its strong points, you should avoid any reform in existence for five or more years that is not able to meet these C-level research-and-practice specifications.

3. Practical Considerations

Before selecting a program, think about whether you have the resources necessary to make it succeed. If there isn't enough money to do the program right or if the administration, faculty, and community associated with a school don't have the will to make the program work, you should not attempt it. Failure is terribly expensive, in terms of dollar costs and professional morale, and it is best to conserve both until you find a program with which you can succeed.

Remember that in a free-market system, the buyer has to beware. Today, the descriptions of most reform programs sound more like marketing campaigns than scientific presentations. Without careful probing on the part of a school, the marketers will not always come forward with a full list of the fiscal or human costs required to make a design succeed. This fact underscores the importance of three types of data gathering I described above: the examination of case studies, phone interviews, and site visits. Here are some of the specific resource issues you should consider:

■ Out-of-pocket dollars. What could the costs be for the program's first three years? For equipment and materials? For training? For teachers' time? How much will it cost to sustain enthusiasm for and commitment to a reform in years four and beyond? (Schools and districts almost invariably underestimate these costs, and the program developers may not be much help.) Where will those dollars come from?

■ Work hours. How many people will have to work how many hours per week for how many months or years to achieve what you expect from the program? If some of this will be unpaid time, which is likely to be the case, will the leadership of the design, district, and school be able to sustain sufficient teacher and administrative enthusiasm to see the design through? Where will the union weigh in on the issue? (Some, but rarely all, teachers view working on a design over the long haul as a benefit rather than a cost.)

■ Adequate technical assistance. What kind of help can you expect from the group providing the program? Will there be enough technical assistance so the people in your

school will understand the program's broad design and get the specific skills they need to make it work? Some programs do not offer any help at all. If you are interested in such a program, ask yourself if it is likely to take hold in your school without outside support. You should know, too, that some programs are growing so rapidly that they are hard-pressed to give the kind of help a school may need. Other programs' offerings, however excellent, are expensive, although they may offer special considerations to districts that involve several schools simultaneously.

■ Missed opportunities. An important intangible is the other opportunities you may be giving up in order to implement this program. Are there equally defensible uses of the school's time? Better ones? If yes, what is the justification for embarking on a project that will consume time, money, and energy?

Finally, most reforms involve some kind of tradeoff. Here are some you should be aware of as you consider a program for your school:

■ Organizational reforms vs. curriculum reforms. If a reform changes school organization—substituting, for example, a traditional organization in which the principal makes all the important school decisions with a school council organization where power is shared with teachers and perhaps parents—this is likely to involve human costs, but it rarely adds financial burdens to a school. Changing curricula is almost always expensive, both in terms of new materials needed and the professional development needed to prepare the teachers to teach the new program.

■ Reforms that involve the entire faculty vs. those that don't. If a reform involves only part of the faculty, training costs less and the people who don't care to participate simply don't—that is the up side. On the down side, such reforms often lead to charges of special treatment and them vs. us attitudes. Also, it is worth remembering that students experience the whole school, not a part. The Obey-Porter amendments, described in the previous article, clearly favor whole-school and presumably, whole-faculty designs.

■ Specified vs. "constructivist" reforms. Reforms that are highly specified—that is, they spell out the curriculum and materials to be used and even teaching strategies in considerable detail—are often perceived as requiring more work and more change. If implemented in any credible fashion, however, a constructivist approach is usually much more work. It often requires the people involved to both create a curriculum and then figure out how to put it into practice. Human and financial costs of a specified approach are typically front-loaded (that is, they are relatively obvious, and many of them come with an up-front price tag). By contrast, costs of a constructivist approach are typically back-loaded (e.g., their human and fiscal costs, which may not be apparent at the beginning, become obvious over time).

Every month, a half-dozen different magazines trumpet the joys and advantages of new reforms. The quotations from enthusiastic teachers and photographs of engaged students can be very seductive. So can the promotional

materials from the developers of these programs. But even after you've discounted the hype, sifted the wheat from the chaff, and found a program that seems well suited to your school and its resources, you need to think hard about whether the people whose effort will make the program work have the will to commit and stay committed to a particular design for at least three to five years. If not, I advise against starting. Reform is hard, slow work. There's no gain to be had, and substantial loss of faculty morale, when you change reforms every few years. The will to stay the course is itself a scarce resource that needs to be carefully allocated.

As the above pages make soberingly clear, the efforts required to choose, implement, and institutionalize a school restructuring design are substantial. The benefits to students and professionals, however, can be equally great. In every profession and every industry, the press in the 1990s is for "continuous improvement." Schools need to be in the forefront of this movement instead of lagging far behind.

Examine your options aggressively. If you choose a reform program, implement your choice relentlessly. And as the old cowboys used to say, "When you get to the end of your rope, tie a knot and hang on!" Education reform makes for a very educational, but sometimes very bumpy, ride. □

REFERENCES

- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 19(12), 997-1003.
- Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Ellis, A.K., & Fouts, J.T. (1997). *Research on educational innovations*. Larchmont, NY: Eye on Education.
- Fashola, T., & Slavin, R. (1997). Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk*, 2(3), 251-308. New York: Harper & Row.
- Popham, W.J., & Sirotnick, K.A. (1974). *Educational statistics*, 2nd ed. New York: Harper & Row.
- Slavin, R.E., Madden, N.A., Dolan, L.J., Wasik, B.A., Ross, W., Smith, L., & Dianda, M. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk*, 1 (1), 41-76.
- Stringfield, S. (1997). Underlying the chaos of factors explaining exemplary U.S. elementary schools: The case for High Reliability Organizations. In T. Townsend, (Ed.) *Restructuring and quality: Problems and possibilities for tomorrow's schools* (pp. 143-160). London: Routledge.
- Stringfield, S., Millsap, M., & Herman, R. (1997). *Special Strategies for Educating Disadvantaged Children: Results and Policy Implications*. Washington, DC: U.S. Department of Education.
- Stringfield, S., Millsap, M., Herman, R., Yoder, N., Brigham, N., Nesselrodt, P., Schaffer, E., Karweit, N., Levin, M., & Stevens, R. (1997). *Special Strategies Studies Final Report*. Washington, DC: U.S. Department of Education.