

Education Week

The Dalmatian and Its Spots

January 28, 2004

By Siegfried Engelmann

At least part of the problem educators have in establishing effective instruction has to do with the illogical recommendations that researchers make. This illogical reasoning occurs in just about all research-based recommendations since 1985, when "Becoming a Nation of Readers" was published.

Why research-based recommendations fail logic 101.

This illogical practice is the confusion about what follows from a true statement. Here's a noneducational example:

*If a dog is a Dalmatian, it has spots.
Therefore, if a dog has spots, it is a Dalmatian.*

The first statement is true. The second statement doesn't follow from the first.

The probable response from most readers is that nobody could be naive enough not to recognize this flaw. English setters, some terriers, sheepdogs, and many mutts have spots. Unfortunately, there are many educational parallels to the argument that all dogs with spots are Dalmatians. Here's one:

If a beginning-reading program is highly effective, it has various features: phonics, phonemic awareness, and so on. Therefore, if a program has these features, it will be highly effective.

Current reform practices revolve around this logic, but the logic is as flawed when it refers to effective programs as it is when it refers to Dalmatians.

Here's how the flawed reasoning occurs. Investigations like that of the 2000 report of the National Reading Panel start by sorting through research studies to identify specific programs that work. Call this group of programs *Dalmatians*.

Next, the investigators analyze the group of Dalmatians to identify their common features. Call each feature a *spot*. They find that the more effective beginning-reading programs have common features (phonics, phonemic awareness, decodable text, oral practice formats, and others). So they have formulated the true statement parallel to: *If a program is a Dalmatian, it has spots*. (If it is an effective program, it has the common features A through N.)

Next, investigators formulate their flawed recommendations, which assert (or imply) that if a program has phonics, phonemic awareness, decodable text, oral practice formats, and so

forth, it will be highly effective. In other words, the investigators' conclusion is parallel to the conclusion, *If a dog has spots, it is a Dalmatian*.

The conclusion has no logical basis. There is a lot more to a Dalmatian than having spots, and a lot more to programs that generate superior outcomes than having the features that are specified in recommendations. The additional features would include the amount of new material introduced on each lesson, the nature of the reviews that children receive, the ways in which the program tests mastery, the number of times something is presented in a structured context before it occurs in other contexts, and many more technical details about how the material is sequenced and field-tested.

But the investigators do not simply flunk Logic 101. They set the stage for a daisy chain of illogic. Because the analysis has removed spots from Dalmatians, they are no longer Dalmatian spots, just spots. So the analysis moves from a more careful articulation of each *Dalmatian* (effective program) to an elaboration of *spots*, now freed from the constraints of the effective program.

Phonemic awareness is a spot. The analysis of the spot goes something like this: "Let's see, there are different types of phonemic-awareness activities. There's oral blending, rhyming, alliteration, segmentation, phoneme insertion, and phoneme deletion. Therefore, any combination of these activity types would meet the requirement of phonemic awareness, and the best versions of phonemic awareness would have all types."

Another problem with 'research based' recommendations is that the investigators apparently do not research the skill and capacities of the consumer of instructional practices.

If researchers conduct experiments to validate their notion of phonemic awareness, they typically don't compare their results with those of a highly effective program in terms of total time required and the performance outcomes. They are satisfied if their intervention results in a gain in performance on some standardized measure.

Note that the illogical formula for the design of programs would create benefits for districts that were using programs that had no spots. A program constructed from spots would probably produce results better than those of the programs the districts are using.

So if a little better is what districts want, that's what the "spots first" reasoning will probably deliver. Unfortunately, the criteria become a double-edged sword that may reject truly effective programs.

The full circle of the daisy chain occurs when a state takes these "research based" recommendations and uses them as adoption criteria for programs that are supposed to be effective, but rejects a true Dalmatian because it does not meet the "standards" the state has set. For instance, a "standard" might indicate that the program had to have the full range of phonemic-awareness exercises (including activities that are ill-suited for beginning at-risk students, like phoneme deletion). If effective program X does not have *all of them*, it fails to meet a "research based" standard, even though it is highly effective and there is no evidence that the adopted programs are effective.

Not only is this type of reasoning possible, it happens with frightening regularity. For instance, California's *Ventura County Star* carried an article on March 15, 2003, titled "Effective Reading Program Must Go." A school in the district, it said, "was the only school in Ventura County and one of 109 in the state to get the citation ... for showing exemplary progress." The district was replacing the program with one that has no strong data of effectiveness, but that had been adopted by California because it meets the state "standards."

The county superintendent justified the move this way: "We want to make sure all schools are using the same curriculum. Why not something based on the standards that are going to be taught?" So in the end, the state not only identifies mutts as Dalmatians, but rejects true Dalmatians because they don't meet the state-created definition of "Dalmatians."

The solution is to excise this medieval logic and to be more straightforward about identifying *specific programs that work*, without pretending that the analysts are able to identify the full set of variables that make the program effective. This is not to say that the criteria for effective instruction are unspecifiable, only that the current standards are far from specifying them, and the effort of trying may be misplaced. If the goal is to identify programs that are effective, why not take the most direct route and simply identify them without the questionable analyses?

Another problem with "research based" recommendations is that the investigators apparently do not research the skill and capacities of the consumer of instructional practices (aside from possible verbal reports). The result is that even if their analysis disclosed all the vital characteristics of effective programs, their recommendations for using the evidence on effective instruction would completely lack research support.

For example, the April 2000 "Report of the National Reading Panel: Teaching Children to Read" discusses phonemic awareness, and the panel makes this recommendation: "There are many ways to teach [phonemic awareness] effectively. In implementing [phonemic-awareness] instruction, teachers need to evaluate the methods they use against measured success in their own students."

The assumptions are that a mix-and-match creation by the typical teacher will be effective, and that the teacher knows how to evaluate the methods he or she uses against measured success. There is no data showing that typical teachers are able to successfully combine components to make superior instruction, and none to suggest that a significant number of them have the knowledge or the resources needed to operate on the implications of "measured success," particularly if they are unaware of what a truly effective program is able to achieve. Before issuing this recommendation, a research-based panel would first have gathered data to address some practical issues:

How many years would it take for an average teacher to "discover" or "create" an excellent combination (given that it would be hard to try out more than one or two combinations a year in a classroom)? What kinds of records would be needed to make this enterprise systematic? How does this pursuit fit in with the district-adopted program and practices? Where does the teacher get the funds and the time that may be necessary to evaluate the results?

Two issues are even more serious: What concern do we have for the children who are being subjected to the teachers' experimentations, particularly if it takes the assiduous teacher years to come up with a program that has sufficient "measured success"? What in the history and demography of teachers in failed schools suggests that more than a very small percentage of them would be able to develop highly effective packages without *extensive* training?

The ultimate products of the National Reading Panel's spots-first logic are implications that true Dalmatians are not really Dalmatians. "[I]t is more common for phonics programs to present a fixed sequence of lessons scheduled from the beginning to the end of the school year," its report says. "In light of this, teachers need to be flexible in their phonics instruction in order to adapt it to individual student needs."

The ultimate products of the National Reading Panel's spots-first logic are implications that true Dalmatians are not really Dalmatians.

The central problem with this appraisal is that to accept it, one would have to deny that Dalmatians are Dalmatians. Highly effective programs have a fixed sequence. When the panel calls for adapting instruction to individual student needs, it is implying that the successful sequences are not successful, and that the teacher will be able to improve on the program by deviating from the program's "fixed sequence."

In fact, the highly successful program has evidence of being successful with the full range of beginning readers. This range comprises great variation in "individual student needs." The panel doesn't have to know how the program does it, but the panel must accept the evidence that the program must have successful procedures for accommodating "the needs of individual students."

Certainly, teachers would have to be trained to use the effective program to achieve individualization, but training would present specific practices that have been demonstrated to be effective and efficient. Teachers would not be encouraged to make changes in the sequence before they were very familiar with the details of the program. The training would show how to group children homogeneously, how to place them appropriately in the sequence. Groups may be started in different parts of the sequence and may be moved through the sequence at different rates, with lower performers repeating some lessons, and higher performers skimming parts of the sequence.

If the program is a Dalmatian, however, it has provisions for placing children, teaching them to mastery, and accelerating their performance. Researchers would learn a great deal about both program design and training if they studied effective programs *carefully* before drawing conclusions about what it takes to be a Dalmatian.

Siegfried Engelmann is a professor of education at the University of Oregon, in Eugene, Ore., and the director of the National Institute for Direct Instruction, located there.