TOO SIMPLE TO FAIL

A Case for Educational Change

R. BARKER BAUSELL, PH.D.

OXFORD
UNIVERSITY PRESS
2011

CONTENTS

INTRODUCTION: OBSOLETE FROM EVERY PERSPECTIVE

Thirty-five students sit facing a single teacher. The teacher has just provided a brief but coherent introduction to a new topic, but one portion of her class couldn't follow what she was saying because they have had too little previous instruction on the subject at hand. Another portion of the class is terminally bored because they had previously learned 90% of everything the teacher said (or will say during the upcoming school year). A third contingent is distracted by two misbehaving boys seated at the rear of the room.

Recognizing these problems, and hoping to reinforce the main points of her lecture, she reseats the two boys on opposite sides of the room and has all the students open their textbooks to read the same page. Unfortunately, the same part of her class who couldn't follow her lecture, along with a significant portion of the students who were distracted, also has trouble reading the textbook. And of course the students who already knew what she was talking about already know everything contained on that particular page in their textbook.

Sensing that something is amiss, the teacher decides to vary her routine a bit and have everyone come to the front of the room and sit on the floor surrounding the chalkboard. Following a few minutes of jostling and confusion, the class then watches a student attempt to solve a math problem based upon what has just been taught and read about (by some). This particular student fails miserably and can't follow the teacher's attempts to help him "discover" his error. The remainder of the class isn't at all interested in this process since some of them would have never made such an

egregious mistake, some of them can't follow the teacher's explanation, and some simply aren't paying attention.

Later, with the students back at their desks, the teacher poses a question to the class on the topic. Some students raise their hand whether they know the answer or not; some wave their arms frantically because they are sure they have the correct answer (or simply want the attention); and everyone else waits for either the correct or the incorrect answer, or pays more attention to the myriad other competing activities that are constantly going on in the classroom, somewhat analogous to a cocktail party in which we stand in a crowded room with sounds and conversations going on all around us and must decide to what we will direct out attention and to what we will only pretend to do so.[1]

What these and most other classroom instructional activities have in common is their mind-boggling inefficiency, the amount of time they consume, and the fact that at any given point in time only a portion of the students involved will be actually attending to them—either because the instruction isn't keyed to their particular needs or they are free to attend to competing activities that they find more interesting. And as if all of this were not enough, the teacher herself is most likely ill trained for her job. She probably graduated from a university-based school of education, which may have been staffed by faculty who knew very little about how to maintain order in a public school classroom, make instruction relevant for as large a percentage of such a classroom as possible, foster learning under typical classroom conditions, or even how to teach the types of content she is now charged with covering. And if teaching children to read is part of our teacher's duties, she may have never even been given a cursory lesson on basic phonics instruction. In fact, it is possible that this teacher may never have enrolled in a single course that actually prepared her to teach children to read, to write, or to understand mathematics—perhaps because *her* faculty were never taught that themselves. An accident of history, perhaps, due to the discipline's early thinkers (such as Herbert Spencer, John Dewey) who were less concerned about increasing the amount students *learned* than they were about the philosophical and social implications of schooling.[2] Or, of later popular theorists such as Jean Piaget, whose work would ultimately wind up having no recognizable application to classroom instruction.

But returning to the 35-student classroom, our intrepid teacher realizes that she can't spend any more time on this particular lesson and must

move on whether everyone is ready or not. She therefore announces a quiz on the topic for the next day and hands out a worksheet that she painstaking constructed herself and assigns it as homework in preparation for the impending quiz. Naturally, by now, she knows that some of her students will complete the worksheet conscientiously and some won't because (a) they don't have the requisite skills; (b) their parents don't have them either, so they can't help their children with the assignment; or (c) there is no one in the children's home who has accepted the role of delivering supplementary instruction or monitoring homework completion, believing instead that these tasks are the school's job.

But the teacher doesn't feel too badly about the job she's doing. One portion of her class is doing quite well, primarily those who listen in class, complete their homework, and whose parents are themselves adequately educated (and consequently recognize the necessity of being involved in their children's education). What our teacher probably doesn't realize is that if an age-appropriate aptitude or cognitive test of any sort had been administered to her students when they were three years old, the resulting scores would have very nicely predicted the identity of the children who do and do not complete their homework assignments—and probably even who will and will not graduate from college.

For, in truth, classroom instruction adds surprisingly little value to the preparation that parents provide their children in the home. True, new topics are introduced and old ones embellished during the 12,000-plus hours that children spend in school, but years before school begins some parents also contribute thousands of instructional hours to their children's education by exposing them to a challenging vocabulary, talking to them about the world, reading to them, instilling in them the importance of learning, limiting their television viewing to educational programming, and teaching them the alphabet, their numbers, word recognition, and often even how to read fluently. After school begins, these same parents also monitor what is going on in the classroom and, if the schools are not meeting their expectation, do not hesitate to intervene by requesting a new teacher, providing supplementary instruction (either themselves or by engaging tutors), ensuring that homework assignments are completed, or sending their children to private schools if necessary. And, not surprisingly, it is the children from these homes who do the best in school and who so please our hypothetical teacher by their performance in her classroom.

It is also the children from these homes who help disguise just how abysmally obsolete the classroom model has become over the years. For it is the presence of such children (and the schools they attend) that allows educators to remain entrenched in their practices and to support business as usual, pointing to the performance of these children and their schools as proof that classroom instruction does indeed work well under the right conditions. Of course, these conditions always involve the presence of students who come from learning-enriched home environments and can read as many words on their first day of school as their counterparts from economically deprived environments (who not coincidentally also happen to be assigned to attend "poorly performing schools") will be able to read by the end of their first *year*. And so what if these poor-performing elementary schools feed even worse performing middle schools and high schools until another generation of adolescents graduates without being able to read a newspaper or write a coherent sentence? At least, our current schools work well for *some* children.

But do they? What if no children arrived on the first day of school with any previous academic instruction? Would this permit the same degree of complacency? Could we afford to tolerate the resulting performance from our obsolete instructional system?

The sad truth is that no one knows just how *little* value classroom instruction adds to children's education, but it is the performances of our inner city schools serving children from home environments providing little or no supplementary instruction that probably give us the best indication. For here, at least, we can see the pathetic results of 12,000 hours of *classroom* instruction delivered to children who do *not* receive thousands of hours of extra-school parental *tutoring*.

But while everyone who knows anything about education knows how important these early home-learning factors are, absolutely *none of us* knows how much sheer human potential is squandered by our continued reliance upon classroom instruction delivered in the form of a poorly equipped, poorly trained teacher standing technologically naked in front of 35 diverse students. No highly educated parent bothers to look at the results of a typical inner city school district and say, "there but by the grace of God goes *my* child." On the other side of the coin, however, few people look at the graduates of one of our well-regarded suburban (or private) high schools and ask, "how much more potential would these

educationally fortunate young people possess if they hadn't been taught in such an obsolete manner?"

We certainly can't rely upon our current testing system to give us any hint about any of this, for our tests are as obsolete as our classrooms. Indeed, our "achievement" tests aren't even designed to assess what is learned in school. Instead, they were developed via an obsolete century-old intelligence testing model designed primarily to rank order students based upon the types of home environments they came from.

One wonders how even the most demented committee conceivable could have designed a more *inefficient* mode of instruction or a more disingenuous method of disguising that inefficiency. Yet, educators seem completely committed to this woefully obsolete model, in part through simple inertia and a desire to avoid the effort that change always entails, in part because we are all so wedded to the concept of one teacher standing in front of a group of students that we are blind to the obvious option staring us in the face. But as recent history has shown, sometimes change is inevitable when it is technologically driven and obviously superior to business as usual.

What I propose to do in this book, therefore, is to explain what science tells us about the direction in which this inevitable change must move. In so doing, I will present the simplest conceivable theory of school learning and the equally simple (but not necessarily obvious) instructional principles that flow from it—all of which has one purpose: to show how our current obsolete mode of classrooms can be transformed into a learning environment capable of dramatically improving the education of *all* society's children. And, while I have chosen to concentrate on elementary school *instruction* because of the crucial importance of mastering the basic academic skills taught there, the principles I will present here are equally applicable to middle and secondary schools as well.

From a personal perspective, this book represents the culmination of an interrupted intellectual journey that began many years ago. It presents the synthesis of the entire field of school-based learning that has simmered uncompleted, like a low-grade irritant in the back of my mind for three decades. A synthesis that ultimately reduces to the preeminent importance of increasing both the *amount* and the *relevance* of the instructional time we provide our children.

My particular journey actually began in 1968, with my enrollment in a doctoral program in the University of Delaware's College of Education and with the subsequently unparalleled, exhilarating opportunity this

provided me to conduct research into the factors influencing classroom learning. From that research, and the work of other researchers who came before and after, I came to realize that something was very wrong with how we herded our children into boxes in order to teach them. I also came to realize that something was very wrong with how we explained—both to ourselves as educators and to the world at large—why some children appeared to learn with such greater ease than others. But, for some reason, it took me a long time to realize that the solution to increasing school learning (as well as to explaining why some children perform so much better on standardized tests than others) boils down to the one simple factor: *relevant instructional time*.

Why it took me so long to fit the pieces together of this exceedingly simple puzzle, I do not know. Perhaps it was due to the happenstance that forced me into a field of research outside of education, thereby distracting me from addressing the puzzle's solution. Or, perhaps it was simply difficult for me to accept the fact that the entire discipline in which I had been trained boiled down to a single elemental concept—*time*—and that everything else proposed to explain school learning was nothing more than a chimera, a proxy for this single variable.

But, for whatever the reasons, hopefully the journey ends here with a completed theory of classroom learning and the crucial (and unavoidable) implications it provides for guiding us to exponentially increase school learning. For, after all of these years, I would not bother with the effort if I did not believe that we *now* have the technological capability for not only improving school learning, but also for eliminating the educational disparities that our obsolete classroom methods accentuate. Nicholas Lemann perhaps best articulates the problem when he says that this country has "channeled opportunity through the educational system and then . . . failed to create schools . . . that would work for *everybod*y, because that was very expensive and voters didn't want to pay for it."[3] To which I would add: "nor did educators have any idea *how* to create such schools."

As will become clear in the chapters that follow, however, education is an exceedingly simple discipline—far more so than anyone realizes—thus, it follows that any theory emanating from it capable of solving our schools' inadequacies must be simple as well. Educational research is equally straightforward, so while I will briefly discuss a few of my own experiments (as well as some truly seminal work conducted by others) to illustrate the

scientific basis for the direction we must take, this book isn't really about research, science, *or* theory. It is about how we can solve one of the most bedeviling problems facing us as a society: how to make the schooling process more productive for *all* of our children.

Of course, anyone who follows educational issues over the years knows that there has been no lack of opinions regarding how we should reform our schools. I will even touch on some of the more promising of these, although most have no scientific basis and stop short of informing us about anything that will actually impact *learning*.

Fortunately, however, both the theories and the research that informed them, share one characteristic: Unlike the more sophisticated sciences that must employ complicated mathematical formulations or complex neurobiological processes to explain their principles, everything associated with education and school learning is exceedingly *simple*. As an author, this provides me with a huge advantage, for not only do I have uncomplicated subject matter to discuss, my audience is exceedingly knowledgeable and experienced, having spent a significant portion of their lives receiving classroom instruction.

But, just because something is simple does not mean that it is either self-evident or unimportant. The success or failure of our schools has far-reaching implications, not only for the children who attend them but for everyone with an interest in the well-being and future of our society. Parents, because they entrust their children to the schools to prepare them for a future increasingly dependent upon knowledge and the ability to apply it; society because, at this very moment, we may well have a potential Newton, Darwin, Gandhi, Shakespeare, Mozart, or Einstein spending her childhood moving from one obsolete classroom to another in an inner-city school. And no one anywhere can believe that such a child could realize even a fraction of her enormous potential in such a place.

So, my sole intent in this book is to enumerate principles and strategies to increase school-based *learning*. I recognize the importance of philosophical, social, and political issues involved in the educational process, and I realize that the schools exist for purposes in addition to the production of learning.[4] I will, however, leave these larger societal issues to those wiser than I, and deal with my limited area of expertise: learning. I wouldn't even hazard a guess as to how we can produce future scientific, artistic, or social leaders, such as the luminaries just mentioned. What this

book deals with is making sure that *all* of our children have the opportunity to learn to (a) read fluently, (b) write coherently, and (c) apply mathematical concepts in their lives. It is also very much about providing *all* of our children with the opportunity to realize their ultimate potential for contributing to our society and maximizing their chances for attaining a high quality of life therein.

CHAPTER 1

The Science of Learning

To improve learning, we must first understand what it is. Although scientists are beginning to make exciting inroads into identifying the chemical and biological changes that occur in the brain during the learning process, we are light years away from being able to apply any of their findings to classroom instruction. But fortunately, from a behavioral (as opposed to a biological) perspective, learning has been the subject of serious study for the past century, and although some of this research occurred in one of the most artificial learning settings imaginable—laboratories employing both animals and undergraduate psychology students—even this work has generated principles that have direct applicability to optimizing classroom learning.[1]

CLASSIC LEARNING RESEARCH

Ultimately, learning entails a neurobiological response to a stimulus of some sort. This unobserved neurological response is translated to an observable behavioral response, which can encompass anything from avoiding the stimulus in the future to correctly answering a test item. Classic learning research (as well as educational research in general) primarily concerns itself with changes in such behavioral responses (*learning*) following the presentation of visual or oral stimuli (*instruction*). Thus, if students are able to correctly answer test questions following instruction that they couldn't answer correctly beforehand, then we *infer* that learning has occurred.

Just as all learning basically involves some type of observable/measurable behavioral response, instruction also always boils down to a stimulus that is capable of eliciting such a response. From this perspective, then, instruction can take the form of (but is not limited to) such diverse stimuli as:

- Being lectured to in a classroom setting
- Completing computerized/online instructional modules
- Being presented a word, phrase, or nonsense syllable and told to memorize it
- Completing homework
- Engaging in self-study
- Reading
- Being read to
- Watching television
- Surfing the internet
- Listening to others (whether in class or at the dinner table)
- Being the beneficiaries of direct parental teaching
- Being corrected by parents
- Observing and subsequently modeling parental or peer group behaviors
- Observing the environment
- Visiting institutions with instructional agendas such as a churches, museums, and science centers

To control as many factors as possible in their research and to avoid teaching something that their subjects had already learned, classic learning studies often employed the visual presentation of nonsense syllables via a technique called *paired-associate learning trials*. Experimental subjects (typically, college undergraduates) were taught, via repeated presentations—often involving a slide projector or its equivalent—to "pair" these syllables (or sometimes conceptually unrelated words) until this arbitrary association was successfully "learned." To avoid as much error as possible in inferring that learning had occurred (and to measure it as precisely as humanly possible), testing involved exactly the same processes that were used in instruction (i.e., the syllables, words, or whatever, presented via the same medium in which they were learned).

As obsolete as current classroom instruction is, present-day teaching isn't quite this rote. Still, unlike classroom research, these experiments

employed a form of instruction and a method of measuring learning that could be controlled and repeated quite consistently. This meant that scientists could have a great deal of confidence in any learning principles they unearthed. Whether these principles would apply to all types of learning, no one knew for sure, but the best guess was (and is) that the same neurobiological processes are associated with all types of learning resulting from all types of instruction, rote or creative, interesting or dull.

So, at the risk of oversimplification, three facets of learning were inferred by these studies, based on how many trials (or how quickly) students mastered the paired-associate tasks for which they received "instruction." These learning facets or parameters were:

- *Original learning,* which is identical to what we mean when we refer to school learning;
- *Retention,* which refers to how long what was learned is remembered—or to the circumstances under which forgetting occurs; and
- *Transfer of learning,* which in classic learning theory refers to the fact that previous learning can sometimes facilitate (and sometimes even impede) subsequent learning.

And, if you think about it, these three behaviors pretty much reflect what we expect students to take from the schooling process: learning what is taught (otherwise attending school is a total waste of time), remembering what is taught (because if we don't remember what we've learned, we might as well have not learned it in the first place), and being able to apply what is learned to new situations (because supplying correct responses to test items would be worthless if we can't assume that this will ultimately be related to other types of innovative, creative, or compliant behaviors of societal importance).

In a nutshell, then, the principles emanating from this type of research that were most relevant to classroom instruction and student learning were:

1. The more times the paired-associate tasks were repeated (that is, the more *instructional time* supplied), the more learning occurred. This was the strongest and most consistent relationship that this line of investigation ever uncovered: more relevant time on task (or more presentations of the stimuli) results in more learning. It was so pervasive, in fact, that some researchers embraced a "total-time

hypothesis," which basically postulated that, within reasonable limits, the same amount will be learned in a given amount of time regardless of the number of trials presented within that time period.[2]

2. Some forgetting almost always occurs, but the more time on task (or the more presentations of the stimuli), the longer the association (or learning) was retained (remembered). Retention can also be improved by (a) increasing the meaningfulness (or relevance) of the content and/or (b) continuing to present the stimuli even after they are learned (which was called *over-learning*). Of course, this still reduces to time on task (or increased instructional time) since the presentation of a stimulus is a form of instruction.
3. Transfer of learning (one form of which was called "learning to learn") proved to be a more tenuous affair, but it does occur as a function of instruction under certain conditions. For example, transfer was facilitated by over-learning, and it occurred most reliably when the training conditions were most similar to the ultimate testing conditions (which in schooling terms is reflected by practices such as teaching to the test or teaching test-taking skills) and when the original learning task possessed certain components in common with the transfer task (such as teaching a child the sound representing a certain vowel to facilitate the learning of a word containing that vowel). However, we still haven't learned enough about this concept to stretch it to what we mean by such attributes as creativity (or innovativeness), and this remains a major gap in our understanding of the instructional–learning process. Suffice it to say that the occurrence of learning is a prerequisite for both retention (and transferring that learned knowledge to novel applications), but learning is no guarantor of either.[3]

Now, admittedly, this brief overview does not do justice to classic learning research. Other variables were involved[4] but, generally speaking, most of the work in classical learning research, as in educational research in general, never transcended what educational researchers in my day called the "grandmother principle," which can be summed up in the following succinct generalization:

You never discover anything in educational research that your grandmother didn't already know.

Still, our grandmothers weren't always right about everything, so it doesn't hurt to subject some of their opinions to scientific tests. Thus, in summary, far and away the most important finding emanating from this classic research (as well as from learning research that involved rats navigating mazes) was that the strongest determinant of laboratory learning is the *amount* of instruction delivered. More instruction, more learning; more time spent studying, more learning; more time on task, more learning; the more time an author spends repeating something, the more likely the reader is to learn it—to remember it—and to apply it.

CLASSIC SCHOOLING RESEARCH

Understandably, researchers interested in studying classroom instruction couldn't help questioning the broader relevance of the classic laboratory investigations of undergraduates paid to memorize nonsense syllables. They felt a need to study children actually being taught in a classroom setting. Thus, they tended to do their research based upon what real teachers did with real students within real classrooms.

In so doing, these researchers both gained and lost something. What they gained was the ability to observe learning in the real-life school settings in which they were primarily interested and to which they aspired to generalize their research. What they lost was any real degree of control over the research setting, in the sense that they had to deal with (a) much more diverse students who, unlike the undergraduates participating in paired-associate experiments, could not always read or understand directions; (b) teachers who potentially could vary in their instructional ability and conscientiousness; and (c) tests that weren't designed to match what students were taught (i.e., standardized achievement measures).

Still, some of this research, much of it conducted before the field's steroidal boosts in the mid to late 1970s—which I attribute to (a) Gene Glass' popularization of meta-analysis[5] (that, among other things, definitively demonstrated the positive learning effects of small class size[6]) and (b) Benjamin Bloom's emergence as the preeminent learning theorists/researcher of the 1970s and 1980s[7]—did uncover some very interesting findings, even if none quite transcended the "grandmother principle."

Some of the more important of these findings as they relate to school learning included:

Increased Instructional Time (or Time-on-Task)

Despite the obvious differences in settings, the classic learning principle that more instructional time (although classic learning researchers seldom labeled their presentation of nonsense syllables *instruction*) results in greater learning did indeed apply to the classroom. In its most elemental form, the more time that is allocated to teach a topic, the more students will learn.[8] In fact, the amount of instructional exposure is one of the strongest determinants of school learning yet discovered.[9]

Of course none of this would come as a surprise to anyone's grandmother. Neither would secondary evidence showing that children who are assigned homework (which, after all, translates to extra time-on-task) learn more than those who do not[10] or that those who attend summer school (which involves increased instructional time) learn more (or forget less) than those who do not.[11] Other similarly obvious manifestations of the relationship between instructional time and learning include the negative impact of school absences and even tardiness.[12]

Strangely, given its obvious importance, as far as I'm aware no one made a serious attempt to document the dose–response relationship between the amount of school instruction until the mid-1970s, when David Wiley and Annegret Harnischfeger[13] conducted a secondary analysis of data from 40 Detroit schools contained in the Equality of Educational Opportunity Survey. Defining the number of hours of schooling delivered to students in any given school, they used the following simple formula:

[# Hours of Instruction Delivered = Daily Attendance (which encompasses absences) x # Hours in the School Day x # days in the School Year]

They found huge discrepancies in the total number of hours of schooling in this one city, ranging from 710 to 1,150 hours per year. "Typical pupils in some schools receive 50% more schooling than pupils in other schools." Then, controlling for student characteristics as best they could, they found that "over a year's period … in schools where students receive 24% more schooling, they will increase their average gain in reading comprehension by two-thirds and their gains in mathematics and verbal skills

by more than one-third" (p. 9). Needless to say, this finding reflects an *extremely* powerful relationship between the *amount* of school instruction and student learning.

Yet, as powerful a factor as the amount of instructional time is, historically it has not been found to be the most powerful determinant factor influencing school learning. That distinction belongs to a relationship that was probably recognized the first time children were ever grouped together in classrooms.

Individual Differences Between Children

Based upon a number of studies (primarily involving large test score databases), it has been estimated that from 40% to 60% of all the individual differences in later school achievement can be predicted as early as the fourth year of life. The best known of these studies was conducted by James Coleman, a sociologist whose 1966 report ("The Equality of Educational Opportunity") definitively demonstrated that *the most powerful determinants of success in school lies in what children bring to the schooling process, rather than what happens to them once they get there.*[14] This is also reflected by the fact that standardized tests administered to children at age three are strongly predictive of test scores obtained throughout their schooling experience.[15]

In a nutshell, what these studies demonstrate (and there are a plethora of them), involving different databases such as the National Longitudinal Survey of Youth and the National Assessment of Education Progress and different types of tests,[16] is that:

- The higher the parents' educational attainment and income level (which reduces to socioeconomic status), the higher the children's achievement.[17]
- Caucasian and Asian students perform significantly better on standardized tests and on just about every other indicator of schooling success than black and Hispanic students.[18] (Of course, race and ethnicity are also related to socioeconomic status.)
- Children from single-parent homes (and especially those in which the mother is very young) fare worse in school.[19] (This also is related to socioeconomic status and race, since 70% of black children are born to single mothers.)[20]

- Children with many siblings[21] do more poorly on standardized tests. The spacing of siblings (closer together is detrimental because of less time available for the parent to interact with any one child) and birth order are also important for the same reason.[22]
- Students who are the beneficiaries of a home-learning environment characterized by (a) plentiful reading material,[23] (b) procedures to restrict the type and amount of television viewing and video game playing,[24] and (c) parents who read to them when they were young achieve significantly higher than children who come from homes without these advantages.[25]
- Children who are actively taught the alphabet, the sounds letters make, words, numbers, number concepts, and even how to read prior to attending school obviously do better in school than do children who are not so taught.[26]

Historically, there has been a great deal of disagreement among educators and educational researchers over the question of why some children seem destined to succeed in school and others seem destined to fail. Some have seen these findings as irrefutably supportive of the heritability and preeminent importance of intelligence, aptitude, and/or ability, whereas others have visualized them as primarily environmentally determined. As will be discussed in Chapter 4, however, these findings possess a considerably more parsimonious explanation.

Instructional Methods

So far, we've only discussed one school-based *intervention* that has any positive effect upon school learning, and that is the amount of instruction delivered. Children who are given more instruction learn more than those who are given less. Surely a more mundane finding is difficult to envision.

Unfortunately, although researchers have evaluated just about every other factor imaginable, not much else appears to influence school learning. Every so often, however, someone comes along and recommends this or that instructional method—such as the use of visual aides, hands-on activities, certain types of discussion groups, discovery learning, educational games, or some other combination of bells and whistles—based upon the belief that his or her brainchild should produce superior learning.

Intuitively, this is quite appealing, for even our grandmothers would agree that the way in which children are taught ought to make a difference in how much they learn. And, at a tautologically absurd level, this is certainly true, such as delivering a lecture to non-Asian American students in Mandarin versus English.

But alas, whenever a sane innovative method is compared to the same amount of traditional classroom instruction, the result is always the same. No statistically significant difference. One method is just about as effective (or ineffective) as another *as long as the amount of instructional time is controlled*.

There are two important caveats to this statement, however: First, if the new approach involves teaching a different subject or a new set of skills to the exclusion of something else, then obviously students will learn more of the new subject (or set of skills) than will students who weren't taught it, *if* the test used to evaluate the new approach measures this new material. (This is a combination of classic time-on-task and common sense.) Also, if the new approach involves teaching prerequisite skills not taught via the traditional method, then the former will most likely be superior to the latter if (*and only if*) these skills are sufficiently useful (and, of course, the test is appropriate). The best example of this is the inclusion of a phonics component in reading instruction. If one group of students is taught to read phonetically by learning to sound out the syllables of words and another group is taught to read by learning words by sight (i.e., memorization), then even if instructional time is controlled, the students taught to decode the phonetic structure of words usually learn to read faster.[27] There is nothing that earthshaking about this phenomenon. It is comparable to saying that students who have mastered algebra will learn calculus faster than those who have not, because calculus employs algebraic constructions hence prior instruction in algebra translates to *additional* instruction in calculus. The second caveat involves interventions that increase the *relevance* of the instruction delivered to the learner because this has the effect of increasing *time on task* (which is the same thing as increasing instructional time). Examples involve not teaching content the learner already knows (which would obviously make the instruction irrelevant regardless of how much of it was delivered) and reducing classroom distractions (which would require more instructional time to produce the same degree of learning). Both strategies are enhanced by reducing class size and (most notably) by tutoring, but let's save these latter issues for

later and use the remainder of this chapter to discuss the preeminent role of instructional time in determining the amount children learn in school.

Methods Versus Programs

The equivalence of different instructional methods should not be confused with different programs of instruction. Contemporary examples of the latter are listed in the Institute of Education Science's "What Works Clearinghouse." Usually, when such programs report positive results, a closer examination will determine that they (a) entail extra instructional time (in comparison to their control group) and/or (b) their content is more closely matched with the standardized tests used to assess student learning.

An excellent example of one of the high-quality trials appearing on the IES website is a study entitled *The Enhanced Reading Opportunities Study*[28] in which 34 high schools from ten districts were randomly assigned[29] to either receive the program or not. The program basically involved 225 minutes of literacy instruction *on top of* the students' regular ninth-grade language art classes (obviously a huge increase in instructional time). The experimental high schools were further randomly assigned to receive one of two different instructional methods. The results were that the *experimental program* resulted in significantly superior reading comprehension skills for those students who received it than for those who did not. However, there was no difference between the two instructional methods comprising the program itself (because both received the same amount of additional instructional time), although of course both were superior to the control group (because its students received significantly *less* instruction).

School and administrative restructuring

To a certain extent inspired by the *No Child Left Behind* (NCLB) legislation (which constituted a bizarre attempt to legislate *school learning*)[30], there have been a number of administrative (e.g. district wide reforms based upon corporate accountability models) and school restructuring (e.g., breaking up large urban high schools into smaller ones—primarily championed and funded by the Bill and Melinda Gates Foundation) initiatives

in recent years. School districts have also experimented with outsourcing the management of their schools to for profit corporations as well as various school choice initiatives (most notably the charter school movement). In general the results emanating from evaluations of these interventions have been uniformly disappointing, although most of this research is so poorly controlled as to be scientifically meaningless[31]. Diane Ravitch, a well regarded educational policy expert, provides a thorough narrative review of this research in her very informative and readable book entitled *The Death and Life of the Great American School System* [32]. Once a vocal supporter of both NCLB and many of the accountability/school choice initiatives, Dr. Ravitch later changed her position while still managing to provide the most even handed historical perspective on these issues of which I'm familiar.

Aptitude-by-Treatment Interactions

Historically, the absence of research pointing to the superiority of any instructional methods over others was completely counterintuitive to many educators. There just *had* to be some instructional methods that would dramatically increase student learning in the schools! Surely, there were some methods of instruction superior to simply standing in front of a class and teaching! After all, don't we all have different learning *styles*? Don't some people prefer visual versus auditory presentations of information or more participatory methods, for example?

Well, we may have different learning styles, and some people may *prefer* one method of instruction over another, but this particular attribute (or preference) doesn't appear to affect learning one iota. Nowhere is this better illustrated than in the case of a well-known educational psychologist named Lee J. Cronbach, who in the late 1950s gave a stirring call to arms on the topic in his inaugural presidential address to the American Psychological Association.[33]

Dr. Cronbach advanced a deceptively simple (and intuitively attractive) hypothesis for explaining why nothing seemed to work better than anything else in the classroom. Turning the concept of learning *styles* on its side, he suggested that it was their ubiquitous *presence* and potency that explained why there seemed to be no difference between teaching methods (and presumably why most educational innovations didn't seem to work to advance learning).

Professor Cronbach hypothesized that, in a research study contrasting a new, well-conceived innovation such as Instructional Method X with an old standby such as Instructional Method Y, there would surely be a significant cadre of students (with, say Attribute A, whatever "A" happened to be) who would benefit from New Method X but who would actually learn *less* when taught by Traditional Method Y. Unfortunately, there would likewise be another cadre of students with, say, Attribute B, for whom the opposite would be true. They would learn more when taught by Traditional Method Y but less when taught by New Method X. Thus, when the two Methods were contrasted with one another in the same research study, the learning styles of the two types of students would cancel each other out, thereby disguising the fact that there really are very important differences between the methods.

Soon published in an article titled "*Two Disciplines of Scientific Psychology,*" this paper generated a great deal of excitement among educational researchers because it explained the frustrating plethora of studies resulting in "no statistically significant difference" that had characterized schooling research for decades. Dr. Cronbach went on to call for a research initiative designed specifically to identify those "aptitudes" (which included not only learning preferences but also such student characteristics as ability, gender, and ethnicity) that conspired to mask the effectiveness of the interventions designed by our best and brightest educators.

The proposed existence of these hypothesized "aptitude-by-treatment interactions" was especially attractive to schooling researchers, who were beginning to realize that they were members of a failed discipline in which absolutely nothing worked better than anything else to increase learning. (With the ubiquitous and powerful exception of increasing the amount of instruction delivered—but since everyone's grandmother already knew that more instruction was better than less instruction, this didn't count, and this relationship was often ignored.)

Yet, despite the hypothesis' promise, it had one small problem. No one could find these dueling attributes. Even worse, a thorough review of the research literature by Glenn Bracht, an educational researcher, conducted a few years after Professor Cronbach's clarion call, basically concluded that the techniques for identifying these effects "was often an afterthought rather than a carefully planned part of the experiment" and that "this approach has not been successful in finding meaningful disordinal[34]

interactions" (p. 639). In other words, such effects were not factors in either schooling research or schooling practice in 1970, and alas nothing has intervened in the ensuing decades to change that conclusion.[35] Incredibly, Cronbach himself later acknowledged researchers' failure to find his cherished interactions but, undaunted, suggested the abandonment of statistics and science in favor of "intensive local observation" since "too narrow an identification with science … has fixed our eyes upon an inappropriate goal."[36]

Fortunately, this tenacity in the face of overwhelming negative evidence has not harmed Lee J. Cronbach's scientific legacy, and he is remembered for more memorable achievements. As far as the science of schooling is concerned, however, the unfortunate bottom line is that research on "learning styles," like research contrasting different ways of teaching, has been an exercise in futility. Neither is a serious factor in classroom learning.

Another Caveat

Obviously, everyone knows that some types of students learn more (or more quickly) from instruction than others. It is therefore not impossible to find "ordinal" aptitude by treatment interactions involving differences in "ability level" (or amount of prior knowledge) in which, say, high-ability students learn more from one type of instruction (or all types of instruction) than do low-ability students. What is difficult (if not impossible) to find is a method of instruction that benefits one type of student but not another when both types of students have the necessary prerequisites for learning the content being taught.

Teacher Differences

But surely, schooling researchers reasoned, if individual differences among students constitute the most potent determinant of school learning, then individual differences in teachers must also be an important factor in classroom learning. Common sense would seem to tell us that this *should* be the case, since we've all experienced both good and bad teachers during our schooling careers, even though we're usually judging them on qualities other than their ability to elicit higher test scores. Perhaps one teacher seemed to particularly value us and/or our potentials. Or, perhaps some had a gift for enlivening their classes with humor or interesting asides or unwavering enthusiasm for an otherwise boring subject. So, although we

all personally probably know what good teaching means to us personally, the sad truth is that educational researchers, despite myriad attempts, have been unable to consistently identify teachers who, year-in-and-year-out, produce superior *standardized test scores* than their peers.

There are several reasons for this difficulty. One is the questionable propriety of employing standardized tests primarily to rank order students on their knowledge of certain relatively ill-defined subject matter content followed by a subsequent *re-ranking* of teachers based upon the same data. (We'll discuss some of the deficiencies of standardized tests in more detail in Chapter 8.) Another problem is that test scores are influenced by so many factors other than teachers, such as differences in (a) children's home learning environments (which include direct parental instruction, parentally instilled expectations for achievement accompanied by incentives/disincentives far more effective than anything a teacher can bring to bear in a classroom, supervision of homework/study assignments) and (b) classroom ambiances (e.g., the need to constantly discipline disruptive students or the presence of extremely heterogeneous students with different instructional needs).

We also don't have a particularly strong theory for why two identically trained individuals with identical amounts of experience standing in front of identical classrooms and teaching the same topic for the same length of time *should* produce different results, unless one of the instructors:

- Had a communication deficit that prevented students from understanding him or her (which hopefully is quite rare among teachers) and/or
- Couldn't maintain sufficient discipline to ensure that his or her students were attending to the instruction (which is possible, but chances are that such a teacher would eventually either learn certain rudimentary class management skills or leave the profession).

Of course, given the causal relationship between instructional time and student learning, we would predict that if some instructors devoted a higher proportion of their classroom time to actual instruction than others, then their students would be expected to learn more. (And, as will be discussed shortly, there is indeed research indicating that major differences do exist among teachers with respect to how *much* time they

actually devote to instruction.) We also know that if some instructors teach material more closely aligned with the end-of-year standardized test, then *their* students will perform better in those tests.

Unfortunately, until recently, there had been very little research to indicate whether teacher differences, if they exist, are consistent from year-to-year. (Obviously, even if we could identify teachers who are effective one year but *ineffective* the next, the information would avail us nothing.) And, although a limited amount of research has attempted to ascertain if teacher behaviors in general are stable across time (based upon the assumption that if teachers don't teach in a consistent manner from year-to-year, then their student learning probably won't be stable either), the results of this line of work have been generally negative.[37]

True, in the past there have been several studies that demonstrate modest teacher effects[38] upon student learning, but most of this work was fatally flawed because it didn't follow teachers longitudinally, nor did it adequately take the huge individual differences among students' propensities to learn into account. Some studies control for little more than the proportion of students in each school who receive federal lunch subsidies, arguing that once this is done, any systematic differences in test scores between *classrooms* must be due to teacher differences. After all, what else could it be?

Well, I'm sure that just about everyone can come up with a plethora of alternative explanations, such as students' past academic performance, the possibility that some teachers are systematically assigned children with poorer (or superior) educational prognoses, and so forth. But even those studies that do attempt to take these factors into consideration seldom attempted to assess the consistency of teacher performance. So, although some studies that have employed large student/teacher/school test score databases have shown that students taught by more-knowledgeable teachers (or teachers who are certified[39]) achieve higher test scores than those of less-qualified teachers, it is also true that suburban schools are able to attract better-qualified teachers than are impoverished inner-city districts.[40] And, assuming that achievement differences as dramatic as those that occur between children from, say, professional families and single-mother welfare recipients can be statistically subtracted out by simply controlling for factors such as racial mix or the proportion of students receiving free lunches borders on the absurd.

A truism, law, or educational fact of life is that no statistical procedure can make an apple an orange, nor can *anything* control for socioeconomic learning differences when it isn't the socioeconomic differences themselves that *cause* these learning differences. The real factors that *cause* childhood differences in learning, which just happen to be associated with socioeconomic factors (hence ethnicity and poverty), are children's home learning environments and their parents' behaviors.[41]

Now, obviously, no one really believes that *some* teachers aren't better than others, or that some teachers don't devote more class time to academic affairs than others, or aren't more conscientious in covering the curriculum, or don't have a better grasp of the subject matter they are charged with teaching, or can't explain their subject matter better than others. Our problem, as will be discussed in Chapter 8, has been that the huge sets of test scores of questionable validity (that is, that don't actually assess what is taught in any given classroom) have so much accompanying extraneous error (noise) associated with them that they aren't really appropriate for identifying teachers whose students perform consistently better or worse over time. This is not to say, however, that there haven't been some Herculean (and promising) efforts undertaken in this arena.

Value-Added Teacher Assessment

Most commonly associated with William B. Sanders and his colleagues (originally at the University of Tennessee and now at the SAS Institute), one such approach is predicated on the proposition that if enough data on individual students are available over time, then this information can be used to predict these students' test score *gains* in the future. It therefore follows that, if all of any given teacher's students' test score gains can be predicted based upon these students' past performance, then any discrepancies from these predictions represent that teacher's effectiveness-ineffectiveness for that particular year.

Called *value-added teacher assessment*, this approach uses sophisticated longitudinal statistical modeling procedures to generate predictions regarding students' test score gains for a given year. It then defines any observed classroom performance that turns out to be better than predicted on the end-of-year test as the *value added* by the teacher of said classroom. (Again, what else could it be?) This approach has resulted in

some relatively promising findings, especially for mathematics, to a lesser extent for reading, but apparently not so much for other subjects. Before considering these findings in any detail, however, it is worth noting that the model attempts to simulate the situation in which:

- Students are randomly assigned to teachers (which would help to decrease the individual differences in students' propensity to learn between teachers' classes that occur when students are assigned on the *basis* of their likelihood to gain more or less highly on standardized tests—such as occurs when parents request that their children be assigned to a given teacher based upon that teacher's reputation *or* when a principal assigns students that he or she believes will prosper more with one teacher than another *or* when students are grouped/tracked based upon their ability level);
- Students are tested twice per year, once at the beginning of the year and once at the end (because the learning and forgetting that goes on during the summer is not under the control of the next year's teacher but obviously affects how much children improve from the previous May's testing to the next May's testing—which in turn *is* used to judge that teacher's effectiveness);
- Subtract the two test scores for each teacher to get a measure of how much his or her students learned during the year;
- Repeat the entire process the next year;
- Compare each teachers' learning results across the two years after statistically controlling for as many factors not under the teachers' control as possible (such as the amount of instruction students' had previously received, and continued to receive, from their home learning environments).

Since these conditions are extremely difficult to implement (and information regarding children's actual home learning environment is nonexistent) in the real world of schooling, Sanders and colleagues have made a valiant attempt to do the best they can with what is available to them. Their results have generated a great deal of excitement outside education (both President Obama and Malcolm Gladwell are huge fans), but unfortunately, although the value added researchers' efforts are interpreted as showing that teacher effects are considerable in any given year, the results assessing the consistency of these effects over time are considerably less impressive.

In the largest analysis addressing the consistency of his effects of which I am aware, Sanders[42] compared 4906 teachers who remained in the same school three years in a row and who were categorized (using his value-added approach) as producing below average, average, and above average effects. I have taken the liberty of doing my own representation of those results in Table 1.1 below.

Altogether there were 941 teachers who were considered below average the first year, but less than half of these (404 or 43%) were judged to be below average the third year. (Data weren't presented for what happened during the second year.) And remarkably, 111 (or 12%) of these supposed below average teachers were actually judged to be above average in the third year while 45% moved up to the average category.

Of the 1,253 teachers judged to be above average the first year, 136 (or 11%) were actually *below* average the third year and 44% had regressed to the middle category. This left only 45% of original "high performing" teachers in the above average category both years.

Now think what would have happened if the below average teachers had all been dismissed and replaced based upon their first year performances. In 57% of the cases, the schools in question would have lost a teacher who would have performed at an average or above average level two years later. Similarly, if the high performing teachers been rewarded monetarily based upon their first year performance, in over half of the cases (55%) the schools would have wasted their money because these "high performers" had slipped back into mediocrity (or worse).

Table 1.1. The Value-Added Consistency of Teacher Performance

Teacher Performance	Below Average (Year 3)	Average (Year 3)	Above Average (Year 3)	Total of Year 1 Value-Added Categories
Below Average (Year 1)	**43%**	45%	12%	941 (100%)
Average (Year 1)	21%	**59%**	21%	3712 (100%)
Above Average (Year 1)	11%	44%	**45%**	1253 (100%)

To me, the bottom line here is that only in the case of average teachers did the value-added predictive scheme produce a consistency rate of over 50% (as indicated in the bolded percentages in Table 1.1). For below average and above average teachers the consistency of the technique was only 43% and 45% respectively. This level of consistency is much too low to base important policy decisions upon and it is too low to have any true practical implications for improving public school education.

Another large scale analysis involving the consistency of value-added teacher assessment was conducted using Chicago high school ninth-grade math scores and produced similarly discouraging results.[43] Here, only 33% of the teachers found to be in the lowest quarter of teaching effectiveness one year (based upon their students' predicted scores) were also found to be in the lowest quarter the following year (and 35% of this lowest group were actually judged to be *above average* the next year). And, using the same data base, while 41% of the teachers in the top quarter were able to repeat their performance the next year, 36% were found to be below average. Although these (and the previous) results were statistically significant, it is difficult to see how they possess any practical significance whatever. Certainly everyone would be exceedingly disappointed if we bused thousands of high-performing teachers into the inner city to increase learning there, only to discover that over a third performed below average once they got there—thereby validating Yogi Berra's observation that "prediction is very hard, especially about the future."

Still, even though I think everyone who cares about schooling research would love to have a method to predict which teachers will and will not facilitate salutary student learning, I'm afraid that value-added teacher assessment may not be quite what we're looking for. Allow me to illustrate via the following cautionary notes:

Cautionary Note #1. The most serious problem bedeviling the use of test data to evaluate teachers is the very real likelihood that students are purposefully assigned to certain teachers based upon their past test performance (such as honoring parental requests that their high-achieving children be placed with an unusually effective teacher, which in turn would help perpetuate a self-fulfilling prophesy). If this occurs with any frequency, it could completely invalidate the entire underpinnings of the technique. One researcher, Jesse Rothstein, actually attempted to test the effects of this potential nonequivalent student–teacher assignment process using as

close a variant of Sanders' value-added approach as possible.[41] Incredibly, what he found was that the value-added *fifth*-grade teacher effectiveness scores also predicted the same students' *fourth*-grade teacher effectiveness scores quite nicely. Since students' fifth-grade teachers couldn't have possibly had a *causal* influence upon their fourth-grade teachers' effectiveness, something *had* to be very wrong here. Rothstein interpreted his results as indicating that there was something quite purposeful and consistent about the way students were assigned to teachers at the beginning of the year. Of course, another possibility is that there may be something very wrong with the value-added teacher evaluation model itself.

A similarly troubling finding from the Chicago high school analysis just discussed, of which Sanders was an author, was that the value-added effects for English teachers tended to predict their students' math teachers' effectiveness as well. This sounds suspiciously like a glitch of some sort in the predictive scheme itself, although, as is their wont, Sanders and his colleagues put a happy face on this finding, calling it a "robustness check"—whatever that means.

The real question, of course, is why should having an effective ninth-grade English teacher *cause* students to have an effective ninth-grade math teacher? (Naturally, we wouldn't be surprised if English test scores are correlated with math test scores, but the value-added model supposedly controls for this.) Or, stated another way: Why should what students learn in ninth-grade English have a *causal* effect upon what they learn in ninth-grade math? If this occurred in, say, third grade, we could hypothesize that the children's reading improvement helped them read their math textbooks better (or their standardized math test's word problems), but in general most ninth-grade English teachers don't teach basic reading skills or how to facilitate comprehension of math word problems. (Of course, there is no question that many ninth-grade students would benefit from such instruction.)

So, in the presence of nonrandom student assignment to teachers and these backward (Rothstein's work) and sideways (English teachers predicting math teachers' effectiveness) predictive findings, shouldn't we worry just a little about the circularity of the fact that teacher effectiveness and student improvement are based upon exactly the same data (student test scores)? Would it really be so bizarre to speculate about which causes which? Couldn't the students' performance also be conceptualized as at

least partially causing some of their teachers to *appear* more effective (or ineffective) than they really were? To me, this makes more sense than students' fifth-grade teachers' performance "predicting" the same students' fourth-grade teachers' performance.

Cautionary Note #2. In effect, value-added teacher evaluations control what they can and then assume that everything that can't be predicted on the basis of previous test scores must be due to the teacher. (At present, we have no way to disaggregate classroom contextual effects from teacher effects.) This is a rather tenuous assumption, because undoubtedly classroom dynamics play into how much is learned in a classroom over an entire year. Perhaps, unbeknownst to the teacher (or outside of her or his control), bullying is occurring during recess, in the bathroom, or at lunch. Or, perhaps the classroom instruction itself is impeded by an unusually large number of disruptive influences, or the actual physical environment of the room itself is substandard for some reason.

As another example of the dangers of relegating everything that isn't controlled to teacher influences, the fact that tests are administered once a year by necessity assigns any summer learning losses due to forgetting (as is typical among students from depressed home learning environments) to the next year's teacher. This also relegates any new learning occurring over the summer (due to formal or informal summer instruction which is more typical of children from families of higher socioeconomic strata) to the next year's teacher. Perhaps we can eventually develop a method by which these problems can be statistically controlled (possibly by something as simple as testing students at the beginning of the year as well as in May), *but, in the meantime, it is almost 100% certain that uncontrolled home environmental variables overestimate the size of current value-added teacher effects.* It probably also explains value-added proponents' counterintuitive conclusion that teacher effects are more powerful than individual differences between children. (That is, they are ascribing a substantial portion of these differences between children to differences between teachers.)

Cautionary Note #3. There is a school of educational research, to which value-added proponents are charter members, that believes that, if enough data are available, all future occurrences can be predicted with extreme

accuracy (and the effects of all previously occurring causal factors can be whisked away). The problem with most existing educational databases, however, is that they (a) are fraught with error, (b) contain a great deal of missing data due to student absences/family movements, and (c) lack key information on potentially important variables (because the databases were constructed for completely different purposes in the first place).

These limitations in our existing data almost surely *reduce* our ability to statistically control for what is far and away the most potent determinant of school learning: *individual differences among students* and therefore erroneously inflate the effects attributed to teachers. Thus, to the extent to which errors, lack of data, and unknown determinants of learning impede our ability to adjust for these differences, value-added teacher differences will be overestimated because teachers are credited with the outcomes they haven't affected (or with uncontrolled effects having nothing to do with teacher performance).

On the other hand, there is no question that value-added analysts have earnestly endeavored to produce the most accurate predictions for students' performance possible (based upon their past performance) and for this they deserve a great deal of credit. There are situations, however, in which statistical adjustment just can't solve the problems of unmeasured influences on learning. One involves comparing students enrolled (or the teachers who instruct them) in schools serving economically depressed families to those enrolled (or teaching) in schools serving economically/educationally advantaged families. Disadvantaged students most likely will exhibit cumulatively decelerating achievement trajectories as a function of time and exposure to these nonconducive learning environments whereas, in contrast, advantaged students will exhibit increasingly propitious educational prognoses. There is no way that I know of to disaggregated teacher effects from these diametrically opposed learning trajectories because *they occur during the same time interval* and because they will be *more* pronounced each subsequent year than they were the year before.

With all of this said, sometime in the future, a value-added approach to estimating teacher effects may prove workable. Unfortunately, claims for its present validity, characterized by some of its proponents breathlessly positive claims[45] and reluctance to make their work sufficiently transparent to permit independent replication[46] has led at least one inveterate

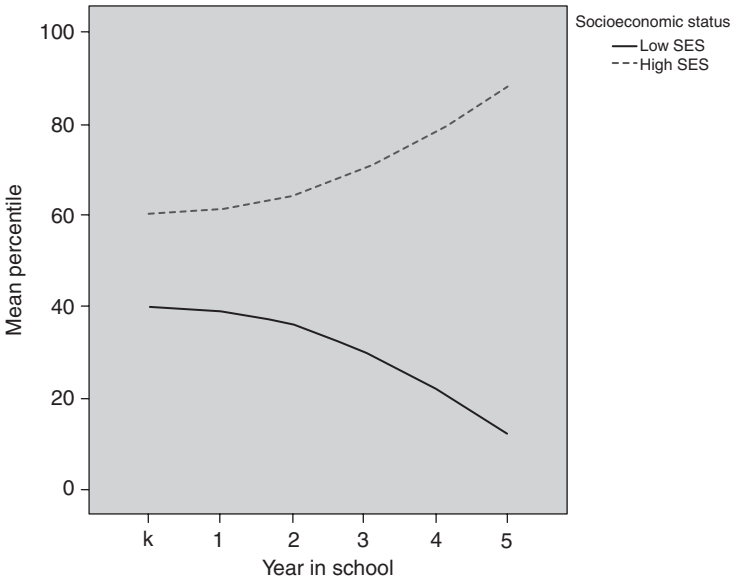

Figure 1.1 Natural Trajectories of Students from High vs. Low Learning-enriched Environments

champion of using student achievement to evaluate teaching performance (W. James Popham, a former president of the American Educational Research Association) to characterize this particular version of value-added teacher assessment as follows:

> There is an old saying that "data gathered with a rake should not be analyzed with a microscope." I think that in Tennessee the rake-collected data are being analyzed with a *mystery* microscope. (p. 270)[47]

So, although we may all believe it is possible (and wish it were already a reality) to evaluate teachers using existing standardized test scores, I'm afraid that value-added teacher assessment really isn't anywhere near ready for prime time. As it happens, however, in the chapters that follow, I will propose some strategies that could greatly increase teachers' ability to produce the learning gains we aspire for all of our children, while at the same time decreasing *differences* between teachers in their ability to do so. But first, let's see what classic research tells us about about what could *cause* some teachers to be more effective than others.

Beyond Value-Added Teacher Assessment

One of the problems with the value-added approach to teacher assessment, which is probably also one reason for its failure to identify teachers who are *consistently* effective or ineffective across time, is its black box approach to the entire process. In other words it employs a strictly statistical strategy for differentiating between teachers without attempting to explain *why* the students of some teachers seem to learn more than the students of other teachers.

As it happens, however, we already know *why*. The explanation is found in the truly seminal piece of educational research called the "Beginning Teacher Evaluation Study" which will be discussed in more detail in Chapter 2. Employing intensive, repeated observations of 25 second- and 25 fifth-grade classrooms, this study found that, on average, 2 hours and 15 minutes of the second-grade school day was devoted to academic activities (which were defined as instruction in reading, mathematics, science, and social studies), whereas 55 minutes was devoted to nonacademic activities (such as music and art), and 44 minutes was "wasted" on things such as waiting for assignments and conducting class business.

Taking math and reading as the two primary academic subjects of interest, the researchers found that, on average, the 25 second-grade teachers allocated 2 hours and 6 minutes per day to instruction. Their students were actually engaged in learning for 1 hour and 30 minutes (or 71% of the time). What was even more telling, however, was the fact that the top 10% (approximately) of the teachers allocated 50 minutes more to instruction than did the bottom 10%, and their students were actually engaged in learning these subjects for about the same amount of extra time (50 minutes).[48] Although this may not sound like a great deal, it means that, in these two crucial subjects, some children could receive 150 hours more instruction during a school year than other students. And, since the average amount of time actually allocated to teaching these subjects was 2 hours and 6 minutes, this means that some children received 71.4 days more instruction than others, or *a total of over 14 weeks of extra schooling*!

To put all of this in context, the investigators contrast two hypothetically average students, one of whom (Student A) receives a grand total of 4 minutes per day of relevant instruction and one (Student B) who receives

52 minutes. Since these students are average, they would start the year at the 50th percentile on the standardized tests, yet by midyear Student A would decline to the 39th percentile, while Student B would improve to the 66th percentile! The authors go on to justify the feasibility of their analyses as follows:

> It may appear that this range from 4 to 52 minutes per day is unrealistically large. However, these times actually occurred in the classes in the study. Furthermore, it is easy to image how either 4 to 52 minutes of reading instruction per day might come about. If 50 minutes of reading instruction per day is allocated to a student (Student A) who pays attention a third of the time, and one-fourth of the students' reading time is at a high level of success [these authors defined "a high level of success" as instruction administered at an appropriate level of difficulty], the student will experience only about 4 minutes of engaged reading at a high success level. Similarly, if 100 minutes per day is allocated to reading for a student (Student B) who pays attention 85 percent of the time, at a high level of success for almost two-thirds of that time, then she/he will experience 52 minutes of Academic Learning Time per day. (p. 23)[49]

So, the moral here is that *massive* differences exist in both the amount of instruction that different teachers deliver, as well as in the amount of *relevant* instruction students *receive*. (We've already mentioned some work[50] that found that the variability in the amount of instruction received by typical students on a school wide basis can be as much as 50%, which borders upon a criminal offense in my opinion.)

So while I haven't seen these studies even mentioned in the value-added literature, in my opinion they constitute the only theoretical rationale of which I am aware for why we *should* be able to differentiate teachers who produce more learning from those who produce less of it. And by simply monitoring classroom instruction by continuously recording it on digital cameras (assuming that provisions were made for constantly identifying opportunities for improvement and then providing sufficient professional development to show teachers how to teach more intensely) we could go a very long way toward either reducing teacher differences in performance or weeding out those teachers who *consistently* teach less. At the very least we could combine these data with value-added procedures,

which in turn might improve the latter's present woeful ability to identify teacher differences that were consistent over time.

Teacher Training

Of course, it could be argued that it isn't even necessary to attempt to properly differentiate between good and bad teachers, given the exemplary training all of our teachers receive. Said another way, perhaps our teacher-preparatory institutions ensure that all of their graduates perform competently, hence negating the possibility of documenting learning production differences among teachers.

One of the first schooling experiments I ever conducted was, in fact, an indirect test of this proposition. My study was inspired by a very famous educational researcher at the time, W. James Popham (mentioned previously as a critic of value-added teacher assessment), who conducted a series of experiments that were designed to find a way to measure teaching proficiency but inadvertently found instead that *neither* teacher experience nor training had any effect upon student learning.[51]

The rationale for his studies was innocuous enough. Popham hypothesized that perhaps one reason we cannot differentiate exemplary teachers from abysmally ineffective ones (always defined, incidentally, by how much their students *learned*) was that our standardized tests simply weren't sensitive enough to measure teacher performance. Just as today, these large amorphous tests weren't that closely matched to the school curriculum, so commercial tests themselves didn't necessarily assess what teachers actually taught in their classrooms. How then could they be used to measure teaching performance? Especially since up to 60% of these test scores are due to individual differences in student backgrounds, thereby leaving only 40% to be explained by other factors (of which teacher differences may account for only a small percentage).

So, Popham decided to start from scratch and develop a series of *teaching performance tests*. First, he designed experimental units based upon discrete instructional objectives, which reflect small pieces of instruction that can be tested directly such as:

> *Sample Instructional Objective:* "Given any two single digit numbers, the student will be able to supply their sum."

Then, each instructional objective was accompanied by a test item that assessed its mastery:

Sample Test Item Assessing this Objective: 7 + 4 = ___.)

The use of instructional objectives and tests based upon them accomplished two crucial functions:

1. They ensured that the teachers knew exactly what they were expected to teach, and
2. The resulting tests assessed exactly what the teacher was expected to teach, nothing more and nothing less.

Thus, for our exceedingly simple illustrative instructional objective above (Popham used more complex ones in his studies involving high school students), there are exactly 100 (and only 100) test items that can be generated to assess the degree to which students mastered the objective (and presumably how well the teacher performed her or his job). Before advocating the use of his tests as a full-blown measure of teacher proficiency, however, Popham wisely decided to validate his approach via a technique called the "known-groups" approach.

The logic behind this technique involved finding two groups of teachers who were "known" to differ on the "thing" being assessed, having them teach the same instructional unit to a comparable classroom, and then seeing if the students taught by the two groups differed in the amount they learned. In this case, the "thing" was teacher proficiency in eliciting learning, so the first task was to find two groups of teachers: one of whom was known to be much more proficient than the other. But therein lay a classic Catch 22. How could anyone identify proficient versus nonproficient teachers if a test didn't yet exist that was capable of rank ordering instructional success?

No problem for Popham. He simply defined his proficient group as professionally trained, credentialed, experienced teachers and his nonproficient group as individuals who had never had any formal teacher training or teaching experience, such as housewives, electricians, and auto mechanics. (The housewives taught social studies, while the other two groups taught topics in their respective vocations.)

And, intuitively, how could anyone construct two more disparate groups of instructors than trained, experienced teachers versus untrained,

inexperienced nonteachers? Thus, Popham had done everything he could to stack the experimental deck (which is appropriate in this instance) to ensure his obtaining huge learning differences between the two groups of students that these teachers and nonteachers taught. So, then this researcher did what all researchers must finally do. He ran the studies and analyzed the results.

While I don't know for sure, I suspect that Popham considered the outcome to be a slam dunk. After all, the experiments' sole purpose was simply to provide a gross validation of a very carefully constructed teacher proficiency examinations (which, in turn, were to be simply based upon how much students *learned* of what they had been taught). And it is worthwhile to note that Dr. Popham was and is one of our most renowned testing experts.

But as the Scottish poet Robert Burns warned us a couple of centuries ago, "The best-laid schemes o' mice an 'men gang (*often*) aft (*go*) agley (*astray*)." The tests functioned quite well for everything except the one purpose for which they were developed. They didn't differentiate between (a) trained, experienced teachers and (b) untrained, inexperienced nonteachers. The conclusion was obvious, if unstated at the time: perhaps (just perhaps) there *really wasn't any difference between trained, experienced teachers and untrained, inexperienced nonteachers as far as student learning is concerned*.

But, although the conclusion was obvious, it wasn't one that I was willing to accept at the time, even though this investigator had basically conducted three separate experiments and found the same thing in each. I was a graduate of a baccalaureate teacher preparatory program, after all, and although none of my courses ever taught me anything about how to teach reading, language arts, or science, for some reason (probably simple cognitive dissonance) I couldn't bring myself to connect the dots. I reasoned instead that the fault must lie in the way the studies had been conducted: one possibility being that it would be much easier to document an effect for teacher training at the elementary school level than in high school (where these particular studies took place). After all, elementary education graduates take many more education courses than do secondary education graduates.

So, being an inveterate skeptic, I set out—with my collaborator, Dr. William B. Moody (who was in charge of preparing elementary school

mathematics teachers at the University of Delaware)—to prove Popham wrong and demonstrate that trained, experienced teachers were indeed better at eliciting student learning than were untrained, inexperienced nonteachers.[52] We designed an experimental elementary school curriculum based upon a set of very explicit instructional objectives that addressed a few number theory topics that we knew elementary students wouldn't have been already exposed to. We then developed a test based upon those objectives (and only those objectives) and located 15 accredited teachers who were willing to devote a week's instruction to them. We also located 15 undergraduate elementary school of education majors who had not yet enrolled in the College of Education course designed to teach them how to teach mathematics (and who had no formal teaching experience).

Each undergraduate was then randomly assigned to teach a comparable classroom within the same schools that housed the real teachers. (Unfortunately, we couldn't randomly assign the trained teachers because they didn't have the time to travel between schools, but we did make sure that they weren't assigned any of their regular students, since this could have conceivably influenced the results.) Both the undergraduates and the credentialed teachers taught the same instructional objectives for exactly the same amount of time for a week. And, at the end of the week's instruction, all of the elementary school students in all of the 30 classrooms were administered the same test based solely upon the instructional objectives that had been covered.

And, of course, the results were the same as Popham's! There was absolutely no difference, not even a trend toward a difference, between the amount the children learned in the 15 classrooms taught by experienced elementary school teachers and the amount the children learned in the 15 classrooms taught by inexperienced, untrained undergraduates. The conclusion seemed inescapable. Teacher training (and perhaps teaching experience) has no (or very little) effect upon student learning. Therefore, should we be surprised if it is extremely difficult to differentiate effective from ineffective teachers (or very effective from moderately effective teachers)?

As I'll discuss in Chapter 5, I even replicated these results later. By that point, I could no longer ignore what my data were telling me, as exemplified by the concluding paragraph I wrote in an editorial for the premiere educational policy journal (*Phi Delta Kappan*) of the time. A paragraph

which also saves me the bother of explaining why I had to seek employment at somewhere other than a college of education:

> Teacher preparation as provided by colleges of education does not result in increased student achievement. The implications of this conclusion are equally inescapable. If the effect of an institution upon its primary purpose is not robust enough to be detected by existing measuring instruments, then the lives of men should not be much affected by its absence. Therefore, given limited educational resource allocations, should we not abandon teacher education?[53]

But, before I share a couple of the studies that completely changed my vision of how children *should* be educated in our schools, I think it would be informative to examine some alternative visions of how school learning can be improved. For the first of these visions, we will have to go back a few years in time to examine how one educational theorist used the research results we've just discussed to come up with a theory of school learning guaranteed to set anyone's teeth on edge who cares about the education of society's children. Research findings, incidentally, which might succinctly be summarized as follows:

> *When it comes to standardized test scores, be they achievement, aptitude, intelligence, or just about anything else, everything is related to everything else, and performance on one test at one point in a child's life predicts performance on another test at another point. When it comes to steps we can actually take to improve learning within the classroom setting (which involves everything from trying to improve teacher education to tailoring instructional methods to students' learning attributes), nothing seems to work except additional instruction.*

For a couple of opposing views, we'll then fast forward to a time after I had left education, and examine a few theories that were informed by some astonishing findings about the educational process that we've only briefly alluded to.