Academic Learning and Academic Achievement: Correspondence Issues

GARY D. PHYE
Iowa State University

The view from Washington D.C. that is offered by Reyna (see Chapter 2) contains the following statement:

The main conclusion that emerges from this analysis is that if this legislation is to be successful, fundamental changes must be made in the kind of educational research that is conducted and in how colleges and universities prepare prospective researchers, practitioners, policy makers, and other educational decision makers.

This message is consistent with that provided by Whitehurst (2003) in an invited address at the annual meeting of the American Educational Research Association.

In this chapter, I provide a discussion of a fundamental addition to the between-group research designs that are used by many educational intervention researchers. The traditional between-group design has been extensively used to assess differences between or among groups of students receiving differing instructional activities. When these between-group designs are used to assess student achievement, the data are typically
reported in terms of grade level averages in a subject (e.g., mathematics) for a particular school. Typically, this is a yearly assessment that reflects group differences between grade levels. The addition of a within-student design element provides the basis for also assessing educational change and individual differences in academic learning and academic achievement. This combining of a between-group comparison with a within-student comparison produces a mixed design with cross-sectional and longitudinal components.

The mixed design has great practical utility when determining the impact in schools of “scientifically based educational interventions.” This utility stems from the fact that the No Child Left Behind (NCLB) Act mandates that schools track the academic development of individual students across their educational careers. Pragmatically, the development in student academic achievement must be monitored in two ways: (a) within a grade level during an academic year and (b) across grade levels during the academic career in that school. For purposes of this chapter, I am using the term academic learning to identify the change in academic knowledge constructed by individual students during an academic year. I am using the term academic achievement to denote changes in academic knowledge constructed by individual students during an academic career.

However, prior to an extended discussion of research design and alignment issues, I want to share my perspective on the NCLB legislation in terms of three roles that educational researchers may play as this legislation is rolled out and implemented. The following roles are defined by what I see as the three major scientifically based research dimensions that define NCLB: (a) the development of scientifically based pedagogical practices, (b) the determination of school system accountability as measured by adequate yearly progress of student achievement, and (c) the determination of the impact of scientifically based educational interventions on student achievement (defined in terms of causal relationships).

**ROLES FOR EDUCATIONAL RESEARCHERS**

The obvious role and the one most frequently addressed in this volume is the educational researcher as the developer of “scientifically based instructional practices.” As has been articulated by several chapter authors, issues such as random assignment, control groups, and clinical trials are major topics that meet the “gold standard” when designing educational research studies.

The second role that educational researchers will play involves the issue of school accountability defined in terms of adequate yearly progress. These activities involve working directly with school districts and include such practical accountability issues as defining adequate yearly progress (AYP) in
terms of learner outcome variables. Again, the crux of these efforts is the development of instruments that provide credible data (reliable and valid) that are collected within a data collection paradigm based on experimental methodology. Several of these issues are discussed by D’Agostino (see Chapter 6).

These psychometric efforts typically involve the determination and development of an acceptable benchmark that operationally defines “proficiency.” The accountability effort is then directed at determining the proportion of a cohort (grade level) that meets or exceeds the benchmark every year with the benchmark for that grade being raised in future years. This is a major challenge for educational researchers in that NCLB calls for the “student” as the primary unit of analysis. In addition, these data must be aggregated to classroom, building, district, and state levels of analysis. Further, state-wide data must also be disaggregated by such student variables as socioeconomic status, gender, second language proficiency, and so forth. In other words, these requirements define a focus on individual differences and group differences tracked developmentally across grades. This is a stark contrast to past practices in many states where achievement data have traditionally been analyzed at the classroom level or district level with no tracking of students across grades.

This same point (unit of analysis) can also be made by raising our level of analysis to a comparison of adequate yearly progress of student achievement outcomes across states using National Assessment of Educational Progress (NAEP) data. For the first time ever, data from all 50 states are available. However, because each state has defined “proficiency” in its own unique manner, directly comparing yearly progress across states is difficult.

The third role that educational researchers may fill is providing evidence that scientifically based practices (role one) used as educational interventions actually affect student achievement. This is an issue of the alignment of learning assessment during the academic year and the end of the year achievement assessment. Here, the educational researcher will be involved with design development, instrument development, data collection, and data analysis that can determine if the observed adequate yearly progress in student achievement can be attributed (said to be caused) by the scientifically based practices that were implemented (as educational interventions) by the classroom teacher. The success of this endeavor will be predicated on the success of the educational researcher engaging in the two previously identified roles to deliver credible evidence about the effectiveness of educational interventions. Given credible evidence of effectiveness, program evaluators may then use experimental research designs to determine the effectiveness of the educational system. These program evaluation efforts will typically focus on the effectiveness of the educational system to impact a teacher’s implementation of the “best practices” and their impact on student achievement when the unit of analysis is the school or district.
This is the scaling up process that has typically been ignored in educational research efforts devoted to systemic change efforts.

Several experimental designs can be developed to address research questions involving instructional impact on student achievement in reading, mathematics, and science within a grade and between grade levels. A central theme common to all such experimental designs using a longitudinal component is the repeated measurement of individual students across time. In effect, this defines a research design that provides the opportunity to assess student change. In the simplest sense, this amounts to viewing academic learning as process (formative assessment) and academic achievement as product (summative assessment) at various points in time.

These research roles provide a schema for identifying research questions that will dictate the development of data collection designs and methods. However, without defining what we mean by academic learning and academic achievement, the communication process that is critical for translating research findings into effective educational practice is hindered. Consequently, the first section of this chapter is devoted to defining academic learning and academic achievement within a theoretical context identified as pragmatism and educational research (Biesta & Burbules, 2003). The second section introduces the use of a mixed research design for data collection efforts that provide a basis for addressing not only achievement status but also individual differences and change in academic learning and achievement.

**EDUCATIONAL RESEARCH: DEWEY AND PRAGMATISM**

**Philosophy of Science**

"Educational research, one might say, is not so much research about education as it is research for education" (Biesta & Burbules, 2003, p. 1). I am couching the following discussions in the context of John Dewey’s pragmatism that shares much in common with that of William James and gave rise to a school of thought frequently referred to in psychology as the "Chicago School of Functionalism" (Bredo, 1997). I subscribe to the position taken by Biesta and Burbules (2003, p. 3): "... we believe that many of Dewey’s ideas are still relevant today—something that, despite the many books that have been written about Dewey over the past two decades, has not yet been sufficiently recognized, at least in the context of educational research." Accordingly, when the question arises about how educational researchers should use pragmatism, Biesta and Burbules offer the following advice:

Pragmatism is not a recipe for educational research and educational researchers; it does not offer prescriptions. It is as we have presented it here, as much a way of
unthinking certain false dichotomies, certain assumptions, certain traditional practices, and ways of doing things, and it can open up new possibilities for thought. It is, in short, a resource that can help educational researchers make their research activities more reflective and—to use one of Dewey's most favorite words—a final time—more intelligent (p. 114).

Pragmatism in the United States has had a pervasive influence in social science research. In addition to John Dewey, early proponents in the United States included such luminaries as William James in psychology, Charles S. Pierce the philosopher-scientist, and the sociologist George H. Mead. Pragmatists argue that philosophy should take into account the methods and insights of modern science (Biesta & Burbules, 2003): “Dewey, for example, stressed the significance of the experimental method of modern science as a model for human problem solving and the acquisition of knowledge” (p. 5).

The main significance of Dewey’s pragmatism for educational research lies in the fact that it provides a different account of knowledge and a different understanding of the way in which human beings can acquire knowledge. Dewey’s approach is different in that he deals with questions of knowledge and the acquisition of knowledge within a framework of action, in fact, a philosophy that takes action as its most basic category. This connection between knowledge and action is especially relevant for those who approach questions about knowledge primarily from a practical angle—such as educators and educational researchers (Biesta & Burbules, 2003).

For both James and Dewey, the basic “function” for human organisms is adaptation to the world. Within this macroview focus (world view), a given is the constant state of change of both the organism and the material world we inhabit. This blooming, buzzing, constantly changing state of our existence as human beings is the norm, not the exception. At this macrolevel of analysis, human beings have the unique ability to cognitively construct alternative hypotheses about “best strategies” for adapting to our evolving world. This knowledge (constructed hypotheses) has no functional utility until action is taken to test the hypotheses. It is the means of testing the success or lack of success of our hypotheses at a macrolevel that characterizes a pragmatic approach to the practice of science in the social sciences.

This philosophical discussion provides a rationale for the use of the term capturing in the title of this chapter. Academic learning and academic achievement are not explanations for change. Rather, they are descriptions of the conditions under which change has been captured. Although this distinction may appear pedantic, “capturing” rather than “discovering” is what educational research is all about. In my opinion, this distinction is the key to successful implementation. Educators must understand that a successful educational intervention that produces positive change elsewhere will be successfully implemented to the extent that comparable conditions exist in their school.
Correspondingly, the microview focus of adaptability can be identified as a single student who is a member of a cooperative learning team or a student engaged in individual seatwork during class. Regardless, the basic tenets of Dewey’s action theory also apply to the analysis of this situation. Students in classrooms are adapting to multiple demands from a constantly changing world. Consequently, their basic function as students is to adapt by constructing hypotheses or plans about the assignment being engaged. These “strategies” are simply ideas that involve symbolic process until they are tested by an action on the part of the student. This action provides the basis for confirmation or disconfirmation of the “plan to adapt.” If successful, the assumption is that this idea would be constructed into personal academic knowledge to be drawn on at a later point in time when similar adaptation efforts are required. In educational circles, this level of analysis of the adaptive function is frequently called “problem solving” and the ability of a student to successfully adapt to successive educational demands to be “problem-solving transfer” (Phye, 2001).

The point is that Dewey’s pragmatic action theory of knowledge construction is predicated on two principles that current experimental psychology as a discipline incorporates into the scientific approach describing, predicting, and explaining human behavior (including cognition). These principles are (a) the rejection of a dualistic approach to explaining psychological reality (e.g., mind/body dualism) and (b) the rejection of a reductionism approach to macro- and microlevels of analysis that insists on “cause-effect” relationships between or among levels with the ultimate explanation being located in the most basic level of analysis. Rather, a correspondence between levels of analysis (Stanovich, 2001) based on both credible and creditable research (see Chapter 1) provide the logical basis for articulating the necessary and sufficient conditions for describing, predicting, and explaining change in student behavior.

Philosophically, Dewey’s pragmatism provides a functional philosophy of human behavior with powerful potential for promoting educational science. I make this statement in light of the continuing debates being waged by educational researchers who continue to resurrect “realism” and “logical positivism” as the straw person against which they rile. Nearly a century ago, Dewey initiated a pragmatic approach to the study of human behavior. These methods of science hypotheses have been tested by the actions of experimental psychology for the last century with remarkable success in promoting our understanding of human behavior. Can we as educational researchers make comparable claims?

The Educational Sciences View

Reyna’s call for a reevaluation of educational research is not a call to “reinvent the wheel” of research methods in human learning and the
assessment of academic achievement. Rather, it is an invitation to examine psychology's contribution to understanding human learning and achievement that has been accumulating at a rapid rate during the past 25 years. Many educational researchers have missed that part of the cognitive revolution in psychology where research methods and statistical tools that produce credible data have been applied in real-world settings. A few examples would include the subdisciplines of developmental psychology, cognitive-social psychology, cognitive-memory psychology, school psychology, and educational psychology. Each of these subdisciplines view the phenomena of learning and achievement through a "different lens." For the most part, however, when the behavior being investigated is academic learning and achievement, three common themes can be observed in the published research: (a) Dewey's theory of action, which amounts to a theory of experimental learning; (b) functional correspondence requiring the use of operational definitions to define actions; and (c) individual differences.

The first theme is the use of an organic model. This simply acknowledges that human behavior is the product of an interaction between a person's environment and his or her genetic makeup; in some respects, everyone is unique. This acknowledgment identifies individual differences as variables that must be considered when describing, predicting, or attempting to understand "how" someone learns or "what level" of achievement he or she has attained. In the research literature, these are commonly referred to as subject variables or abilities.

The second theme is the use of operational definitions when identifying the knowledge to be acquired and/or constructed and the conditions or situation in which this occurred. Is it a group setting or an individual setting? Is the academic material to be mastered math, science, or others? Operational definitions are a critical means of communicating with precision the situation in which a particular person demonstrates the desired learning outcome. In the research literature, these operational definitions of the task and environment are frequently the independent variables to be actively manipulated or passively recognized (field studies) and assessed. These independent variables or fixed variables are operationally defined with precision for a reason. If there has been an impact on the dependent variables (academic learning and academic achievement), strong or weak arguments can be made that the changes in learning or achievement are attributable to the influence of the independent or fixed variables. The utility of operational definitions is their ability to communicate with precision the conditions under which the behavior (action) was observed. In this regard, implementation at other sites (replication) and scaling up at differing levels (classroom, district, statewide, and so forth) becomes a possibility. If the primary focus of a research effort is the validation of best practices within content areas at varying grade levels, followed by classroom implementa-
tion by teachers in various settings (rural/urban), then precision in communication is critical.

The third theme that characterizes much of psychological research is a consideration of individual differences. I am including a consideration of research designs that provide the basis for a consideration of individual differences for two reasons. The first reason is practical. NCLB mandates require the tracking of individual students' academic progress across their academic careers. Further, group differences are also identified as disaggregated data within the larger data set. These groupings include gender, socioeconomic status, language status, developmental status, and so forth. The second reason is theoretical. Both Lee Cronbach (1957) and Benton Underwood (1975) have called for an integration of research design characteristics that would permit the study of both group differences and individual differences, including change across time. Within the context of NCLB research, this translates into designs that provide for the analysis of group and individual student change in the development of academic knowledge (reading, math or science) within the same study.

**Academic Learning**

Learning has so many meanings that it is necessary to first add a qualifying adjective. In the environmental setting we call the classroom, there are different types of learning experiences. Some experiences involve motor learning that serves as the basis for motor skill development. In some cases, learning activities focus on behavior that serves as the basis for social skills development. On the other hand, academic learning involves primarily the processing, construction, and communication of cognitive information (symbolic) about the subject matter we teach in the classroom.

Having identified academic learning in terms of a "type of learning activity," scrutiny of a commonly accepted definition of academic learning serves as a means of considering what is frequently referred to as the "learning process." In the third edition of *Learning Theories: An Educational Perspective*, Dale Schunk offers this suggestion, "Learning is an enduring change in behavior, or in the capacity to behave in a given fashion, which results from practice or other forms of experience" (Shuell, 1986; as cited in Schunk, 2000, p. 2). This definition of learning is consistent with a cognitive focus and captures the criteria most educational professionals consider central to academic learning.

Three elements of Shuell's definition of academic learning (behavioral change, endurance, and practice) deserve further attention, because these elements help define the level of analysis involved in the present use of the term. Together these three elements stress the idea that academic learning engaged in by students is not typically a single instructional experience. It is a process of (a) the acquisition of new information, (b) the refinement and
organization of what is already known, and (c) the successful testing and use of that knowledge. Academic learning is the product of practice (action) that provides the basis for relatively long-term change in one’s personal knowledge. Ideally, this change will increase students’ ability to successfully adapt as they move from grade to grade where the curriculum requires more complex and specialized forms of personal knowledge.

### Academic Achievement

Many distinctive characteristics of achievement tests can be identified. For NCLB, I am mentioning only three characteristics that are salient to our discussion of student learning outcomes as measured by end of the academic year assessments. These assessments may take the form of a commercial standardized achievement test (Iowa Test of Basic Skills) or a state developed test, and so forth.

In either case, these instruments are developed to assess what students have learned from their exposure to classroom instruction (in contrast to aptitude tests). In other words, a primary characteristic of an achievement test is the assessment of prior academic knowledge. With respect to the use of an achievement test to capture academic learning at the end of the academic year, a second characteristic is the static measurement of the status at which a student is performing at a given point in time. Last, but certainly not least, there is no guarantee that the prior knowledge status captured in this single assessment was the product of learning during the academic year. These three characteristics of achievement assessment at the end of an academic year frequently provide the basis for the determination of adequate yearly progress and is typically referred to as a summative assessment (Nitko, 2004).

Used in this way, achievement tests provide an indication of students’ status in subject content areas relative to end of the year grade placement norms or standards. Phrased differently, this is the target behavior at which instructional goals and objectives are directed. An assumption is made that effective instruction engaged in by the teacher during the academic year promotes student learning that will be assessed by the achievement test. When testing this assumption, accountability data will be collected that reflect teachers’ implementation of scientifically based practices. However, these scientifically based practices must be demonstrated to impact the end of the year achievement performance. This is where the problem of correspondence (both theoretically and in practice) becomes a major issue. Stated simply, how does one go about collecting data during the academic year that takes into consideration the hypothesis that academic learning (assessed as process or change) can be demonstrated to “cause” improvement in students’ achievement status measured at the end of the academic year?
EDUCATIONAL INTERVENTIONS AND EXPERIMENTAL DESIGNS

Classroom Example (Data Collection)

How are academic learning and academic achievement alike and how are they different? Pragmatically, one can make the argument that learning and achievement are alike qualitatively, because a researcher is simply capturing the process of adapting to the school environment at different points in time. For example, Figure 9.1 depicts an assessment cycle that might be observed in the classroom. Assume that this is a fourth-grade classroom and the knowledge being assessed is mathematic computation and problem solving. Common sense tells us that students’ learning is assessed in some fashion that culminated in a status check at the end of a 9-week grading cycle when report cards are issued. Obviously, daily and weekly assessments have preceded the status check. Thus, from a data collection point of view, the weekly assessments are said to be formative, and the quarterly assessment is said to be summative.

Moreover, a typical academic year is divided into four quarters with quarterly assessment, with the end of the fourth quarter typically taking the form of a state-mandated achievement test (see Figure 9.2). The assessment scheme in Figure 9.2 is simply an extension of the evaluation scheme portrayed in Figure 9.1. What was identified as a summative assessment in Figure 9.1 becomes a formative assessment in Figure 9.2.

These examples identify the crux of one correspondence issue. Is the knowledge base (cognitive and motivational abilities) being constructed on a daily and weekly basis being reliably assessed at the end of each quarter (content/construct validity)? Further, when the quarterly assessments become predictors of end of the year performance, do they still reflect what is being learned on a daily and weekly basis? Are the cognitive and motivational abilities and skills being constructed by students to meet the demands of quarterly assessments being assessed on the end of the year achievement test?

If not, then one can make the claim that qualitative differences may exist when the summative evaluations from Figure 9.1 become formative.

Quarterly (9 week) data collection cycle

<table>
<thead>
<tr>
<th>Week</th>
<th>W-1</th>
<th>W-2</th>
<th>W-3</th>
<th>W-4</th>
<th>W-5</th>
<th>W-6</th>
<th>W-7</th>
<th>W-8</th>
<th>W-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_7$</td>
<td>$X_9$</td>
<td>$Y_1$</td>
</tr>
</tbody>
</table>

$X = $ Weekly formative assessment
$Y = $ Quarterly comprehensive assessment

FIGURE 9.1
Quarterly classroom data collection scheme
Yearly data collection cycle

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>W_{1}W_{2}Qtr1</th>
<th>W_{3}W_{4}Qtr2</th>
<th>W_{5}W_{6}Qtr3</th>
<th>W_{7}W_{8}End of Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td>X_{1}X_{2}Y_{1}</td>
<td>X_{3}X_{4}Y_{2}</td>
<td>X_{5}X_{6}Y_{3}</td>
<td>X_{7}X_{8}Z_{4}</td>
</tr>
</tbody>
</table>

X = Weekly assessment  
Y = Quarterly assessment  
Z = Annual assessment

**FIGURE 9.2**
Yearly classroom data collection scheme

evaluations in Figure 9.2. In other words, is there a correspondence between the cognitive motivational abilities of individual students being assessed as formative and summative evaluations depicted in Figures 9.1 and 9.2? This correspondence is assumed in many educational settings where the focus is on the development of an aligned curriculum within a grade level. My point, “How do you know”? This qualitative alignment assumption can be tested by taking an individual differences approach to addressing the question. Simply determine the nature of the relationships (correlation coefficients) among the data points within a 9-week quarter (Figure 9.1) and across the four quarters (Figure 9.2).

**Between-Group Data Collection for Adequate Yearly Progress**

Annual achievement testing at most grade levels has become a reality in our public schools. Because this is a relatively new effort, many schools have simply collected achievement data and reported them by groups (grade level) and then compared group performance from one year to the next. Determining the effectiveness (impact) of educational interventions typically takes the form of comparing different groups of students at different points in time.

The following example (Figure 9.3) reflects a simplified between-group data collection procedure involving two groups of students, each assessed at two different times (e.g., this year’s fourth graders assessed in 2004 compared to last year’s fourth graders assessed in 2003). Many schools are using an extended version of this data collection procedure to assess adequate yearly progress as required by NCLB. This is essentially a between-group comparison without a control group (this design actually does have a “control” or “comparison” group, though not a randomized, simultaneous, or good one from an experimental point of view).

Because data are collected from two different groups at two different points in time, when differences are observed in achievement, one can only recognize that a difference was observed. Any observed difference can be attributed to any number of factors because there is no comparison
Grade 4

End of 2003 Academic Year  Baseline Data Collected (Group 1)
Beginning of 2004 Academic Year  Intervention Introduced
End of 2004 Academic Year  Intervention Data Collected (Group 2)

**FIGURE 9.3**
Common annual yearly progress data collection scheme

group (again, there is a comparison group, just not a good one). More important, change cannot be inferred because individual student achievement is not being assessed at two points in time. A typical data collection scheme I have encountered in the schools I work with involves the assessment of academic achievement or adequate yearly progress with a between-groups design that looks something like the data collection scheme shown in Figure 9.3.

For a single grade, the typical comparison is a difference comparison between group 1 (last year’s fourth graders) and group 2 (this year’s fourth graders). However, we have fourth-grade students serving in the baseline condition different from those who are in the fourth grade when the intervention is introduced. Consequently, any observed differences in achievement between 2003 and 2004 can be attributed to individual and group differences in students as well as to the impact of the educational intervention. Alternatively, as a researcher would comment, we have a confounding that precludes us from attributing observed differences solely to the educational intervention.

In most cases, the AYP assessment process involves determining the proportion of fourth-grade students who met or exceeded the proficiency benchmark in 2003 and the proportion of students who met or exceeded the proficiency benchmark in 2004. However, the standard for proficiency increases annually. Thus, the standard for proficiency is a moving target. This is the big concern of educators in the public schools. This is an immediate practical concern for educational researchers producing credible data upon which policy analysis and administrative decisions are based. From an educational sciences perspective, research designs and data collection procedures of the type described in the paragraph above do not provide the credible data-driven evidence needed to change policy or determine the effectiveness of educational interventions (see Chapter 1).

This discussion of the comparison of academic achievement between different groups of students with data collected at different points in time (this year vs. last year) is an introduction to another alignment issue. This type of alignment issue confronts researchers who are trying to determine the success or failure of the NCLB initiatives when the effort is to determine
achievement gains for groups of students within an academic year and across grade levels during successive years.

**Designing a Classroom Experiment with a Mixed Design**

This design or data collection refinement is accomplished by extending the traditional between-group design to include a longitudinal component (repeated measures). This would involve combining the repeated data collection from individual students as portrayed in the short term (Figure 9.1) or the long term (Figure 9.2) with a treatment group and a comparison group. This extension produces a mixed design for data collection by including a within-student factor (repeated measures for each individual). A simple mixed design would be a 2 (group) × 3 (repeated measures) factorial with random assignment of students to a treatment group and a comparison group as shown in Figure 9.4. With students as the units of analysis, the assessment of performance at three points in time (repeated measure) provides the basis for assessing change. For example, with a mixed design we could assess the impact of a supplemental instructional intervention with a group of fourth graders. The first factor, “grade level,” is a between-students factor with two levels (treatment/no treatment) and the second factor, “assessment of the same individuals prior to, during, and following the educational intervention,” is a within-students factor (repeated measure). Because we are tracking individual students across time, we have the basis for describing observed differences in pretest, midpoint, and posttest performance as “change.” The issue here is that we can infer change only when we analyze data for students who have a minimum of two performance scores. Given student mobility, this requires a check for differential dropout rates and so forth. Data collection involving the assessment of academic learning with a mixed design would look something like that shown in Figure 9.4.

<table>
<thead>
<tr>
<th>Educational Intervention</th>
<th>Pretest</th>
<th>During</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (Group 1)</td>
<td>X ------</td>
<td>X ------</td>
<td>X</td>
</tr>
<tr>
<td>Control (Group 2)</td>
<td>X -------</td>
<td>X -------</td>
<td>X</td>
</tr>
</tbody>
</table>

*X = data collection*

**FIGURE 9.4**

Mixed-design classroom data collection scheme
Because this is a classroom learning study and the educational intervention is typically supplementary instruction in a specific subject (math, reading, science), the control group is given a task that requires study but is not from the same content domain as the educational intervention. Assuming regular instruction will promote some improvement in performance across time, the researcher is interested in isolating effects of the educational intervention (supplemental instruction).

Thus, the research hypothesis of interest is the two-way group by repeated-measure interaction. One form of a statistically significant two-way interaction might look something like that plotted in Figure 9.5. To the extent that the improvement in pretest-posttest performance favors group 1, the unique impact of the educational intervention has been determined. Without going into detail, the within-student's factor also increases the power (sensitivity) of the statistical test to detect an intervention effect. In a short-term study, ethical issues can be circumvented by flip-flopping the treatment and control groups following the posttest data collection so that both receive the educational intervention.

A design of this type would typically be carried out at the classroom level in a single school. The unit of analysis is the student not the teacher or type

![FIGURE 9.5](image)

Anticipated two-way interaction for a successful educational intervention when using the mixed-design data collection scheme in Figure 9.4
of intervention. Rather, with random assignment (and the inclusion of multiple-classrooms—see Chapter 1), this design could be used to determine the effectiveness of a scientifically based educational intervention that positively impacts student performance. This would typically be a short-term experiment that would fit nicely within a quarterly reporting interval such as that described in Figure 9.1 and would typically be reported as an instructional-learning study.

**Replication of Intervention**

Now, let us assume that the supplemental mathematics intervention that was tested at the fourth grade in the example above proved to be effective. Consequently, the school administrators made the decision to extend the intervention for the entire academic year. This suggests a replication study with a data collection procedure similar to that in Figure 9.2. This data collection strategy in Figure 9.2 is simply an extension of the data collection strategy introduced in Figure 9.1 and creates the opportunity for a yearlong replication study. However, a constraint is that one cannot ethically withhold an educational intervention from half the class for an entire academic year. In this case, one frequently resorts to a clinical replication with no comparison group (classical one-group longitudinal study) or one develops a quasi-experimental design by creating a “matched classroom” comparison group. Although the development of a matched comparison group involves extra effort, any observed differences at the end of the year can more “possibly” be attributed to the educational intervention. Without the comparison group, one is still confronted with the question “compared to what?” when explaining observed change in achievement during the academic year. See Chapter 1 for alternative research designs for dealing with practical constraints.

**Evaluating a School-Wide Intervention with a Mixed Methods Design**

Further, assuming the intervention was successful at grade four, the efforts could be extended to include additional grades. In this way, the school would be assessing the effectiveness of the educational intervention across grade levels. Extending the math example to include grades four, five, and six would make sense from a curriculum perspective and would provide the basis for a mixed-design data collection example to be used in a program evaluation effort by a single school.

The basic research question is, “Can we take a scientifically based educational intervention that we have successfully demonstrated impacts fourth-grade student achievement and scale up the intervention effort (within the same school), to include grades five and six”? Our example the-
oretically could include grades kindergarten through 12. For purposes of simplicity here, only three grades are considered. In addition, the following example is basically an integration and extension of the data collection procedure in Figure 9.2 and the research design in Figure 9.4.

The following quasi-experimental design involves 3 years of data collection. The data being collected are end of the year achievement test performance for individual students. Data are collected for all students in grades four, five, and six at the end of the current academic year and for the next two academic years (2005 and 2006). At the end of 3 years, the data collected could be organized as shown in Figure 9.6. The Xs simply denote that the state-mandated achievement test was administered to all grades every year. In our example, these data collection procedures produce a data array from which data sets can be developed for testing various research hypotheses about student change (across grades), group differences in student performance (between grades), and teacher/curriculum influence (time-lag comparison).

It must be pointed out that these types of analysis are possible only to the extent that the state-mandated test has been scaled so that comparisons can be made across grade levels using a normalized or standardized score (see Chapter 6).

At the end of the 2004 academic year, research questions about grade level differences can be addressed by making cross-sectional comparisons in achievement scores among grades four, five, and six. This would involve a comparison among groups designated X1, X2, and X3 in Figure 9.6. The research questions addressed would deal with only group differences (developmental differences) because the comparisons involve different students in different grades at a single point in time.

At the end of the 2005 academic year, a between-groups analysis can again be performed to address research questions involving group differences. In essence, this is a cross-sectional replication of the 2004 study.
with different students in different grades. This involves a comparison among groups designated Y1, Y2, and Y3. However, we can now add a longitudinal component to the analysis to analyze for developmental change. This would involve a comparison of data sets X1 and Y2 for students who have taken both tests. Thus, the analysis involves the same students followed across two grades. This comparison is also possible for data sets X2 and Y3. These two longitudinal analyses provide information about developmental change from fourth to fifth grade and from fifth to sixth grade. Further, a comparison can now be made between data sets X1 and Y1 at the fourth grade, X2 and Y2 at the fifth grade, and X3 and Y3 at the sixth grade. These are time-lag comparisons because one is comparing different students at different times for the same grade. Any differences observed would raise questions about teacher or curriculum influences. This time-lag comparison is an extension of the between-groups design identified in Figure 9.3.

At the end of the 2006 academic year, the longitudinal analysis now includes data sets X1, Y2, and Z3. Developmental change can now be analyzed for students who completed grades four, five, and six. Also, the cross-sectional analysis for differences among grade levels can be performed on data sets Z1, Z2, and Z3 as a second replication of the original 2004 analysis to monitor group differences in grade-level performance. Further, the time-lag analysis can be continued by comparing Y1 and Z1 at the fourth grade, Y2 and Z2 at the fifth grade, and Y3 and Z3 at the sixth grade.

The data collection design in Figure 9.6 is a multiple-methods design (Friedrich, 1972; Nesselroade & Reese, 1973). This type of design is of value when research questions address a school's ability to implement a scientifically based educational intervention on a school-wide scale. Granted, this is a quasi-experimental design and there are problems with both the credibility and the generalizability of findings. However, the multiple-methods design is useful for scaling up efforts from a classroom level of analysis to a school-wide level of analysis and can provide useful data to drive administrative decision making and policy analysis at the local school-district level.

**Scaling Up with Clinical Trials**

Theoretically, it is possible to take our example from Figure 9.6 and move it to the next level of replication. However, as pointed out by Levin (see Chapter 1), the tradeoff is one of introducing potentially new problems pertaining to internal validity threats. However, if confronted with the task of scaling up educational interventions that involve a school-wide system, replication with random assignment and appropriate control groups is still the "gold standard." An example would be an educational intervention that takes the form of a new professional development delivery system. Assume
the basic intervention model has been demonstrated to be effective based on an analysis of data collected from a single school (see Figure 9.6). Now, the question is whether or not the successful intervention results can be replicated in other schools.

Without going into detail, assume that the new professional development model that serves as the educational intervention is the pilot project for a statewide initiative. The pilot project data indicate that this new model is a "strong positive" based on 3 years of data collection and analysis. Consequently, a representative sample of schools from within the state are identified and randomly assigned to the treatment group of schools or the control group of schools. In this case, a power analysis was performed to determine the size of the school sample because the unit of analysis is schools (not classrooms or students). Furthermore, our data collection procedure could essentially be the same as that portrayed in Figure 9.6. Again, a 3-year cycle of data collection would be sufficient to provide credible as well as creditable data (see Chapter 1). These data could then serve as the basis for statewide policy analysis and administrative decision making by the state department of education.

**SUMMARY**

My efforts have been directed toward the reintegration of Dewey's pragmatic approach to scientific research into the current discussion of educational science. This theoretical approach has a long history in education and has been successfully used by psychologists to describe, predict, and explain adaptive human behavior (including academic learning and achievement). An organic model of adaptability and change, this approach has a great deal of potential application for educational researchers. The storyline for our discussion of data collection efforts has been hierarchical in nature. The rationale for this organizational scheme is practical and theoretical. Practically speaking, the stakeholders vary as one moves from a consideration of scientifically based instructional practices in the classroom to systems operating at the level of a school district or a state initiative. Theoretically, two obvious issues are involved. The first point is the one for which examples have been given: Research design issues change as research questions change across these authentic levels of analysis. The second point is subtler. This relates to the correspondence principle, when a level of analysis approach is taken to the analysis and interpretation of hierarchical data.

As noted at several points in the chapter, a pragmatic approach to data interpretation can be located on a continuum that reflects the limitations of data interpretation based on the nature of the data collection and analysis. This is typically translated into the concept that data in the social sciences
can be used to describe, predict, or explain the behavior under investigation. This basic theme runs through all chapters in this volume. However, a corollary assumption, the correspondence principle, is frequently ignored. The correspondence principle states that our continuum of data interpretation does not necessarily hold as we move from one level of analysis (classroom) to the next (school district) and design characteristics change. What may have been a data collection procedure that led to a causal explanation at one level of analysis, when used at a differing level of analysis may provide only prediction or description. Simply including random assignment and the addition of a control group to a data collection procedure is a necessary but not sufficient condition for ensuring a credible causal attribution across levels of analysis.

From a pragmatic perspective, scientific methods in educational research are pluralistic. To make significant advances in educational reform, we must ask better questions and use both experimental and quasi-experimental design elements in our data collection efforts. Continuing to view academic learning and achievement as static products of instructional manipulations contrasted between groups is only a first step. Educational researchers must include data collection procedures and data analysis procedures that provide an opportunity to determine who benefits from best practices (individual differences) and how these students develop academically (change) as a result. A limited research repertoire for an aspiring educational researcher is as self-defeating as a limited repertoire for an aspiring musician. As mentors of the next generation of educational researchers, we (graduate faculty in colleges of education) must do better.

References


