

**Table of Levels of Research and Associated Research Strategies and Data**  
Martin Kozloff

A. RESEARCH STRATEGIES
<p style="text-align: center;"><b>1. EXPERIMENTAL RESEARCH</b></p> <p>An experimental research strategy usually involves <b>comparison</b> of two or more situations (classes, schools, districts) that (through random assignment or through matching on important factors, such as race and social class) are the same in every way possible, but that differ in the <b>few factors</b> (such as curriculum materials, teaching methods, leadership style) whose effects or outcomes (e.g., achievement) you are <b>testing</b>.</p> <p>One group is the <b>experimental group</b>, that receives the intervention. The other is the <b>control group</b> (that receives the usual “treatment”) or is an <b>alternative treatment group</b>, that receives a different intervention that is being tested.</p> <p>Certain <b>design elements</b> make the data from experiments more <b>valid</b> (accurate), and therefore more believable and useable. These design elements are:</p> <ul style="list-style-type: none"> <li>a. <b>Samples are representative</b> of the population for which the findings are relevant.</li> <li>b. <b>Randomization</b> or <b>matching</b> are used to produce equivalent groups.</li> <li>c. Variables to measure are <b>clearly defined</b>: conceptual definitions (general) and operational definitions (examples; things to measure).</li> <li>d. A <b>wide range of variables</b> is measured so that a comprehensive picture is obtained; e.g., achievement, student engagement, the quality of a curriculum, the quality of instruction, and teacher satisfaction are all measured.</li> <li>e. The same variables are measured in several different ways (<b>triangulation</b>) to see if they support the same finding; e.g., achievement is measured by a standardized test, by frequent curriculum-based measures (tests that are in the curriculum materials), and teachers’ perceptions of students’ progress.</li> <li>f. Measures (e.g., how to count behavior) and instruments (for tests) are <b>validated</b>.</li> <li>g. Measurement (e.g., testing) is checked for tester or observer <b>reliability</b>.</li> </ul>

- h. Both **pre-tests** and **post-tests** are given. Even better, data are also collected periodically **during an intervention** to see if there is any trend; e.g., gradual vs. rapid change.
- i. The experiment is **replicated** (done again) with **similar samples** (to see if the findings are consistent and strong---not a fluke), and with **different samples** (to see if the findings apply elsewhere).

## 2. SURVEY RESEARCH

Survey means **overview**. The survey research strategy is used to obtain a larger picture.

Unlike an experiment, surveys do **not** involve “**manipulating**” variables; e.g., **introducing** new curriculum materials and seeing if the new materials are followed by (associated or correlated with) changes in student achievement. However, surveys can still be **comparative**. For example, schools that USE different curriculum materials could be surveyed to see if schools using curricula with systematic and explicit science instruction have higher achievement than schools where students are supposed to “**inquire**” and to “**construct** knowledge.”

Several kinds of **survey data** may be collected, but some data are more difficult to obtain than others. For example, a researcher might use **official statistics** from a state board of education to survey teacher retention or student achievement across a state.

Or, a researcher could survey teachers’ proficiency by observing and scoring (on rating scales) the teaching skills of a **large sample** of teachers in a district. However, this would be a lot of work for one person.

## 3. FIELD (ETHNOGRAPHIC) RESEARCH

Field research is done in naturalistic settings (e.g., classrooms). The point is to obtain a big picture of behavior, social interaction patterns, activities, social roles, and physical environment—similar to what anthropologists do.

Field research usually collects both qualitative and quantitative data .

**Qualitative data**---such as persons’ opinions, feelings, explanations, justifications, and categories (such as how persons define a “good teacher”)---are obtained by recording conversations and through informal interviews.

**Quantitative data** might consist of the researcher **counting** what he or she considers important events, such as how often teachers call on diverse learners.

Field research can be used in Level 1 research, to yield findings that a researcher might want to investigate in a more formal way.

Field research can also be used as part of Level 2 and Level 3 research to **complement** research that is experimental or survey. For instance, large scale (Level 3) research might test how well web-based teacher training courses work. Still, there could be **live field observations** in classes to see how students interact with each other and with their internet instructor.

## B. KINDS OF DATA USED IN DIFFERENT RESEARCH STRATEGIES

1. **Tests with quantitative data.** Tests might tell the percentage of problems students solve correctly, or the percentile ranking of students' test scores. Data might be collected by the researcher or from (for example) a district's official statistics.

2. **Structured observations** (e.g., in class) **with quantitative data.** For example, a researcher **counts** specified behaviors or interactions (correct words read per minute), or a researcher uses **rating scales** to measure teaching proficiency:  
The teacher: (1) Often calls on diverse learners. (2) Occasionally calls on diverse learners. (3) Rarely calls on diverse learners.

3. **Field notes.** The observer writes (or audio records) a **narrative** of what is happening. For example, the observer describes (1) the classroom environment (seating, cleanliness, materials, noise level); (2) episodes of teacher-student interaction during instruction; (3) students' engagement (attention, participation); (4) students' interaction with each other.

Field notes may be used in field (ethnographic) research (e.g., on bullying on playgrounds). But field notes could also be used even in experimental research. For example, the research is an experiment to see which form of teacher supervision raises teacher proficiency the most. Field observations could **also** be made of supervision sessions to see first hand how different kinds of supervision are conducted and the immediate responses of supervisors and teachers.

4. **Structured Questionnaires and Interviews.** These consist of a set of questions that are asked of all respondents. The questions can be "closed-ended"---give participants a "fixed" choice. For example,

"Would you say that your principal visits your class to say hello and to encourage students: (a) often; (b) occasionally; (c) rarely?"

You can see that this sort of question (a rating scale) provides **quantitative data**. The researcher later could **calculate the percentage of** "often," "occasionally," and "rarely" responses.

Or the questions can be more "open-ended." For instance,

"Tell me about the methods you use to increase engagement of your diverse learners."

This sort of question yields **qualitative** information. However, the researcher, later, could make a list (and then count) all of the methods that respondents cite. And the researcher could determine how often respondents cited different methods. These would be quantitative **findings** based on qualitative **raw data**.

Structured questionnaires and interviews are often used in survey research, but they could also be used in experimental research to supplement “hard” data on, for example, student achievement measured by tests.

**5. Informal interviews.** These are often used in field (ethnographic) research. The observer asks questions relevant to certain events. Therefore, there is no fixed set of questions; different questions might be asked of different persons.

“How do students end up sitting where they do? For instance, how do some students end up in the front and others in the back of the class?”

### **c. LEVELS OF RESEARCH, WITH EXAMPLES**

Following are examples of research to answer questions. The researchers begin with Level 1 research, and then use Level 1 (tentative) findings to design Level 2 (more rigorous) research, and then use Level 2 findings to design Level 3 (larger scale applied evaluation) research.

#### **RESEARCH QUESTION 1. CAN WE IMPROVE THE ACHIEVEMENT OF DIVERSE LEARNERS?**

##### **Level 1 Research on Improving Achievement of Diverse Learners**

Two years ago, the Sunnyside County School District replaced its largely ineffective elementary grades reading and math curricula with scientifically tested and effective curricula.

The district’s **official statistics** now show that at least 85% of White and/or affluent students in grade 4 read proficiently and pass the state math tests---a big improvement. However, 55% of minority and poor students **STILL** read at the “below basic level” at the end of grade 4; and 60% fail the fourth grade state math test. These students will do poorly in all other subjects; many will drop out of school.

*In other words, teachers now use curricula known to be effective, but most diverse students are still failing.*

The district knows research showing that diverse learners CAN achieve at high levels. Therefore, school failure can’t be explained away by poverty, family background, or culture. For example,

<http://www2.edtrust.org/NR/rdonlyres/5C01C07B-A827-4396-8B9C-9AE3986D322A/0/ilchicagohsprincipals2007.ppt>

[http://www2.edtrust.org/NR/rdonlyres/9D2EBC7D-4424-4A97-812B-9E5D6CAE880B/0/ca\\_haciendalapuente\\_2005.ppt](http://www2.edtrust.org/NR/rdonlyres/9D2EBC7D-4424-4A97-812B-9E5D6CAE880B/0/ca_haciendalapuente_2005.ppt)

<http://www2.edtrust.org/NR/rdonlyres/F88C0659-BFFF-4458-ABAB-225A2F523D6B/0/midetroit2007.ppt>

Administrators want to find out **why** so many diverse learners are doing so poorly in THEIR schools. Maybe this knowledge can be used to foster success. Administrators have specific research question:

1. “What variables are **associated** with achievement vs. lack of achievement?” This is a question for basic---Level 1---research.
2. “If (based on what we learn from Level 1 research) we **remove** variables that appear to weaken achievement, and **increase** variables that appear to foster achievement, will that **improve** achievement in our diverse learners?” This is a question for Level 2 (test the hypothesis in a small setting), and then Level 3 research (test the hypothesis in many settings).

***The researchers don’t know for sure what to look for.*** They are guided by past scientific research that identifies variables that are associated with achievement, but they can’t be certain that this prior research done in other places perfectly applies to Sunnyside. So, they still need to examine achievement in THEIR schools. [Please see the references (at the end of this document) to scientific research relevant to achievement.] Therefore, the researchers begin (Level 1) with a research strategy and data collection methods that provide a **lot of information of different kinds---field research**. [In other words, they cast a broad net, to find just about **anything** that may be relevant to the research question.]

1. The researchers observe in **two schools**:
  - a. One whose population is predominantly diverse learners.
  - b. One that has a small population of diverse learners, but where there are at least five diverse learners in each classroom.
2. Twelve classes are studied in each school--two classes each in grades k-5.

This is NOT a large sample. And it does NOT represent the whole population of schools and diverse learners in Sunnyside County. But it makes no sense at this early point to study a large sample if you don’t know what to look for.

3. The researchers want to see if they get any **clues** about what variables are associated with low vs. higher achievement in diverse learners.
4. Observers **take field notes in each class** one day a week for one hour (varying the time of day) for a school year.

- a. By studying students in the 6 grades (k-5) **at once** (over a year), the observers can see if there are changes in students **within** each grade from the start to the end of the year. They can also see if students are different from one grade level to the next. For example, does achievement go down in the higher grades?
  - b. Because the study is done for a year, it is a **longitudinal** (over time) **study**. Notice, researchers are **NOT** studying the **same group** of students (called a **panel**) as they move together from kindergarten to the end of grade 5. This would take six years. They don't HAVE six years to improve achievement! Instead, they study classes AT each grade level at the same time. **Each class is called a cohort.**
5. Observers are guided by past research that SUGGESTS the importance of certain variables. These are:
  - a. Where different students sit.
  - b. How often the teacher gives students opportunities to respond.
  - c. How often the teacher gives praise and other forms of recognition and acceptance.
  - d. Whether the teacher provides adequate wait time.
  - e. Whether the teacher teaches and prompts students to use important and general pre-skills, such as listening for important information, remembering information and taking notes, vocabulary.
  - f. How patient the teacher seems to be, as judged by facial expression and tone of voice.
  - g. How the teacher responds to off-task or disruptive behavior.
6. The researchers analyze the field notes **within** each grade level. They look for changes in student achievement within each class from the beginning to the end of each year.

The researchers also compare findings **across** the 6 cohort grade levels, to see if, for example, there are changes as students move from kindergarten to grade 5.
7. The researchers find that **there are associations between teacher-student interaction and achievement in diverse learners**. For instance,
  - a. At every grade level (k-5), the majority of diverse learners with **higher** achievement were in classes where teachers made sure diverse learners sat up front; gave lots of opportunities to respond; were patient; gave frequent praise; taught important pre-skills, and responded to off-task and disruptive behavior in a matter-of-fact way.

- b. The majority of diverse learners who were falling behind in each grade level (becoming disengaged, making many errors, learning very little in reading, math, and other subjects), were in classes where the teachers did not teach in the above way.
- c. **At each grade level**, diverse learners in classes where teachers **did not** make sure they sat up front; **did not** give lots of opportunities to respond; were not patient; **did not** give frequent praise; **did not** teach important pre-skills; and **did not** respond to off-task and disruptive behavior in a matter-of-fact way, became **LESS engaged** and made less and less progress from the beginning to the end of the year.
- d. **Comparing the grade cohorts**, researchers found that in the higher grades, diverse learners **started** the year already **less engaged** (low expectations of success?) and with **less background knowledge** needed for the new class, and began to fall behind right away.

The researchers are encouraged that they've discovered a **possible association** between specific features of instruction and student engagement and achievement. However, the sample was small and unrepresentative of the district population. Moreover, **the findings may be a fluke**. For instance, the association between teacher-student interaction and the engagement and achievement of diverse learners, may have been **chance** (and would not be found again), or the association may **not be a CAUSAL one**. That is, maybe something else causes BOTH the teacher behavior and the student behavior. In other words, the researchers **CAN'T be confident** that changing the teacher behavior might have any effect on diverse learners. Therefore, the researchers decide to move to Level 2 research to test these possibilities. [See the section below.]

### Level 2 Research on Improving Achievement of Diverse Learners

The logic in Sunnyside's research is this:

"If certain features of instruction go along with (**predict**) low vs. high achievement in diverse learners, then (if we make sure that teachers use the features that predict success, and remove the features that predict failure) perhaps we can increase the achievement of our diverse learners."

"However, just because one set of variables (inputs) predicts another set of variables (outcomes), does NOT mean that the input variables CAUSE the outcome variables. The only way to TEST whether the input variables are causes, would be to actually CHANGE the input variables and see if the outcomes change."

"In other words, we need to train teachers to do MORE of what is associated with (predicts) high achievement and TEST to see IF in fact the achievement of our diverse learners increases."

“However, we DON’T know if changing the way teachers teach WILL yield high achievement. So, it would be **unethical** to IMPLEMENT changes in teaching for the whole district. Instead, we will do Level 2 research; we will test our Level 2 research hypothesis in a rigorous way, but only in a few classes.”

The Level 2 research hypotheses are these:

1. “If teachers receive a training program that teaches them to change their instruction so that they give diverse learners lots of opportunities to respond (etc.), teachers’ instruction will change.”
2. “If teachers make sure that diverse learners sit near the front; give lots of opportunities to respond; are patient; give frequent praise; teach and prompt diverse learners to use important pre-skills (vocabulary, comprehension, note taking); and respond to off-task and disruptive behavior in a matter-of-fact way---then (given that teachers are using tested and effective curriculum materials) diverse learners will be engaged in instruction and will achieve high proficiency.”

In other words, the researchers have turned their Level 1 findings (about how variables are associated) into hypotheses that tests the APPLICATION of their findings.

Level 1. Found that  $X \rightarrow Y$ .

Level 2. Change X, see if Y changes.

The research must have certain design features in order to obtain findings in which researchers can have confidence.

1. They re-state their research hypotheses as a **null hypothesis**. Why?

Researchers can always find data (cherry picking) to support what they may believe (the research hypothesis). The honest approach is to collect information that COULD show that you are **wrong**. *They null hypothesis states what the researchers would find if their research hypothesis were wrong.*

Research hypothesis. If X changes  $\rightarrow$  Y changes

Null hypothesis. If X changes  $\rightarrow$  Y does NOT change

Here are the null hypotheses.



- a. "If teachers receive a training program that teaches them to change their instruction so that they give diverse learners lots of opportunities to respond (etc.), **teachers' instruction will NOT change.**"
- b. "**If** teachers make sure that diverse learners sit near the front; give lots of opportunities to respond; are patient; give frequent praise; teach and prompt diverse learners to use important pre-skills (vocabulary, comprehension, note taking); and respond to off-task and disruptive behavior in a matter-of-fact way---then (given that teachers are using tested and effective curriculum materials) diverse learners will be **NOT** be engaged in instruction and will **NOT** achieve high proficiency."

Using the null hypotheses, researchers will collect data that **could** show that their research hypotheses are not true.

2. The researchers use the identified features of instruction (in the research hypothesis, from the Level 1 research) to develop **operational definitions** (needed for later measuring) of the teaching behaviors that appear to be associated with student engagement and achievement. Researchers also develop operational definitions of student engagement and students achievement.
3. The researchers **develop a training program** to teach the teachers how to increase their use of the behaviors that were shown in the Level 1 study to be associated with high engagement and achievement; e.g., how to give opportunities to respond; how to respond to disruptive behavior in a matter of fact fashion; how to teach and prompt students to use vocabulary, comprehension, and note-taking skills. The program involves 8 after-school meetings (2 per week for four weeks) and in-class coaching during the four weeks and for two months thereafter. **In other words, the whole study lasts three months.**
4. Researchers develop instruments for measuring the variables.
  - a. For instance, **instruction** will be measured with classroom observations and **rating scales**. For example, the observer rates one teaching behavior like this:

"The teacher firms up and prompts diverse learners to use note-taking, vocabulary, and comprehension skills:

    - (1) Almost every time that these skills are needed.
    - (2) About two out of three times that these skills are needed.
    - (3) About half the time that these skills are needed.
    - (4) About one out of three times, or less, that these skills are needed."
  - b. The researchers also develop rating scales for measuring "**engagement**"---attention, participation, effort and persistence.

"How would you rate students' attention during instruction?

1	2	3	4
_____	_____	_____	_____

Least

→

Most

c. And the researchers measure **achievement** with standardized tests.

4. Observers and testers are **trained** to use the instruments. They watch videotapes and score teacher and student behavior. They practice until **different observers** scoring the **same** tape obtain very similar ratings; that is, they are **reliable**.
5. The researchers identify two schools in Sunnyside County in which to conduct the research. Again, one school is predominantly diverse learners and the other has few diverse learners. Two classes at each grade level---k-5---are selected---24 teachers in all.
6. The 24 teachers are randomly assigned to two groups: (1) experimental group that receives the new training program; (2) control group that continues teaching as they have before.
7. Teachers receive pre-test measurement of teaching skills; students receive pre-test measurement of engagement and of achievement in reading and math.
8. Researchers conduct the training program. Teachers' skills (taught in the program) are measured every two weeks in the classroom using the same instruments as during the pre-test. Students' engagement is also measured at this time.
9. When training is over (three months), the teachers' teaching skills are again measured (**post-test**), to see how far (if at all) they have come from the pre-test. Students' engagement and achievement in reading and math are also measured, to see how far (if at all) the students have come from the pre-test.
10. Researchers **analyze the data** from the experimental group and the control group.
  - a. Assuming that the two groups had similar pre-test scores on teaching skill and student engagement and achievement, how **much** **did each group change from the pre-test**?
  - b. **Did the experimental group improve more** in teaching skill and in student engagement and achievement? If so, where the changes so large that they are not likely to be just chance?

Remember that one of the null hypotheses is that changes in teacher behavior won't make any difference in students. Diverse learners in the control group—with no change in teacher behavior---will do as well as diverse learners in the experimental group.

Let's say that the data show that the null hypotheses is FALSE. They are not supported by the data.

1. The teachers in the experimental group became significantly more skilled with the behaviors taught. The teachers in the control group did not change much at all. In other words, the training program seems effective.

2. Diverse learners in the experimental group improved in engagement and achievement far above their pre-test scores.
3. Diverse learners in the experimental group **decreased** in their engagement; also their achievement was very little above their pre-test scores, and was far below the achievement of students in the experimental group.

In summary, the research hypotheses are supported.

But this is not enough to use the training program in all schools. It might not work with other samples of students, teachers, and schools. Therefore, wider-scale Level 3 research is needed.

### Level 3 Research on Improving Achievement of Diverse Learners

1. Based on their Level 2 research, the researchers develop research hypotheses for Level 3 research.
  - a. “If teachers in **other schools in Sunnyside County** receive a training program that teaches them to change their instruction so that they give diverse learners lots of opportunities to respond (etc.), teachers’ instruction will change.”
  - b. “If teachers in **other schools in Sunnyside County** make sure that diverse learners sit near the front; give lots of opportunities to respond; are patient; give frequent praise; teach and prompt diverse learners to use important pre-skills (vocabulary, comprehension, note taking); and respond to off-task and disruptive behavior in a matter-of-fact way---then (given that teachers are using tested and effective curriculum materials) diverse learners will be engaged in instruction and will achieve high proficiency.”

2. The Level 3 research has most of the same features as the Level 2 research. [Please review these features, above.]

The main difference is that the Level 2 research is **replicated**: (1) with other schools at the **elementary** level; (2) replicated and extended to **middle and high school** levels as well.

Researchers make sure that the **first replication** is with classes that were in the Level 2 **control** groups, so that the teachers and their diverse learners can benefit from what was learned (teacher training works) as soon as possible.

3. Let’s say that the replications at new elementary schools produce the same findings: the training programs change teacher behavior; changes in teacher behavior produce changes in student engagement and achievement. However, the findings are not the same in middle grades and high school classes. Therefore, the researchers must identify differences between the situation in elementary classes and the

situation in middle grades and high school classes. Let's say that a main difference is the much larger gap in the background knowledge of middle and high school diverse learners. Apparently, these gaps cannot be overcome merely by changing some behaviors of the teachers. The researchers hypothesize that ADDING tested and effective REMEDIAL reading and math curricula may help. So, they TEST this hypothesis by providing training programs to middle and high school teachers that also include using remedial materials. These tests show that NOW the changes in classroom instruction DO improve student engagement and achievement.

In summary, the Level 3 research has yielded TWO kinds of effective ways to raise achievement of diverse learners: (1) in elementary grades; (2) in middle and high school grades.

Future research might examine whether improvements in children's engagement and achievement at the elementary grades results in continued engagement and achievement. However, this would be a long (but essential) longitudinal study!

## RESEARCH QUESTION 2. IS ONE FORM OF SUPERVISION (DIRECTIVE VS. NONDIRECTIVE) MORE EFFECTIVE THAN ANOTHER?

### Level 1 Research on the Possible Effects of Directive vs. Nondirective Teacher Supervision.

For instance, a supervisor has a hunch (from earlier observations or from several research articles that she read), that teachers may benefit more from directive supervision, with explicit instruction from the supervisor.

**She does pilot research to see if there is anything to this hunch.** She provides highly directive supervision to five new teachers, and **nondirective** supervision (teachers evaluate themselves and suggest ways to improve their teaching) to five other teachers---to see if there is **any consistent different** in these teachers' improvement.

If the researcher DOES find the expected consistent differences, she next uses Level 2 research to test more **rigorously** (i.e., using design elements cited in the left-hand column) her hypothesis (not merely a hunch) on the effects of directive supervision

### Level 2 Research on the Possible Effects of Directive vs. Nondirective Teacher Supervision.

Let's say that the **researcher finds from Level 1 research** that teachers who received more directive supervision improved instruction more than teachers who received **nondirective** supervision.

The researcher turns this finding into an **hypothesis** to test more rigorously.

“Teachers who receive more directive supervision will improve their instruction and will be more satisfied with supervision than teachers who receive nondirective supervision.” [Notice that several outcome variables are measured, to get a bigger picture of the possible effects of the

The researcher:

- a. Develops a **conceptual definition** of “proficient instruction.” Then she transforms this definition into an **operational definition**---specific features of proficient instruction; e.g., using clear words, using a proper set of examples, correcting errors.
- b. Develops a **conceptual definition** of “teacher satisfaction with supervision.” Then she transforms this definition into an **operational definition**---specific things that teachers might do and say, that represent different degrees of satisfaction with supervision.
- c. Develops **instruments** (See **DATA** in the left-hand column.) for **rating** teachers’ “proficient instruction” and satisfaction with supervision.”
- d. Tests the **reliability** of the instruments to see if different observers score the same teachers’ proficiency and satisfaction the same way.
- e. Develops a **protocol** that tells exactly how to deliver directive vs. nondirective supervision
- f. Trains some supervisors in two suburban schools to supervise in a directive way and others to supervise in a nondirective way.
- g. Randomly assigns **20 teachers** to the two supervision groups.
- h. Pre-tests (rates) teachers’ proficiency at instruction.
- i. Implements the two kinds of supervision.
- j. Post-tests (rates) teachers’
- k. Uses **structured questionnaires** or **interviews** to find out how satisfied teachers were with their kind of supervision.
- l. Analyzes the data to see if the data support the hypothesis; that is, do the data say that teachers who received directive supervision improved more (and were more satisfied with supervision) than teachers who received nondirective supervision?

For example,

What is the average amount of change in skill for each feature of proficient instruction in teachers who received directive vs. nondirective

supervision?

What percentage of teachers who received directive supervision vs. nondirective supervision rated themselves as “highly satisfied”?

### Level 3 Research on the Possible Effects of Directive vs. Nondirective Teacher Supervision.

Let’s say that the researcher finds from Level 2 research that the data support the hypothesis.

“Sure enough, teachers who received more directive supervision DID improve their instruction and WERE more satisfied than teachers who received nondirective supervision.”

Great!

But is one study of 20 teachers in two suburban schools enough to be **confident** that there would be the **same results with different samples** of teachers (older, male, different grades); in different schools (urban, rural) and with different supervisors? No.

Is it ethical for the researcher to say that all schools ought to supervise in a directive way? No.

The researcher can’t be sure that the findings in the first study WEREN’T a chance fluke (that is, wouldn’t happen again if the study could be repeated exactly).

And the researcher can’t be sure that the findings apply anywhere else.

Therefore, the researcher conducts **larger-scale evaluation research**. For instance, the researcher:

a. Does the study again with samples that are similar to the first ones, to see if the results of the first study were a fluke .

[Let’s say there are the same results. Probably not a fluke.]

b. Does the study across the whole district. There are 40 elementary schools.

The schools are **randomly assigned** (by flip of a coin) to the two supervision groups.

A **random sample** of 10 teachers is selected from each school.

Observers are **trained** to use the instruments for measuring proficient instruction and satisfaction with supervision. The **reliability** of their

scorings is checked.

Supervisors in the schools are trained to use the protocols for providing directive vs. nondirective supervision.

Teachers receive **pretests** of their teaching skills; then teachers receive one of the two kinds of supervision; and then teachers receive post-tests of their teaching skill.

Teachers are given structured questionnaires or interviews to find out how satisfied teachers were with their kind of supervision.

The researcher analyzes the data. For example,

What is the average amount of change in skill for each feature of proficient instruction in teachers who received directive vs. nondirective supervision?

What percentage of teachers who received directive supervision vs. nondirective supervision rated themselves as “highly satisfied”?

The researcher finds that, again, directive supervision yields significantly more improvement in teaching skill and significantly higher ratings of satisfaction.

However, the researcher now **disaggregates** the data.

She divides the data on all teachers into groups whose characteristics may be relevant: male/female; veteran/new teachers; ex-military/non military.

Does the teacher’s sex, time as a teacher, or military experience influence the effectiveness of the two forms of supervision?

The researcher finds that in general, directive supervision STILL works best---whether male or female, new or veteran, military experience or not.

However, male teachers and ex-military (males and females) receiving directive supervision **improve more** than female teachers and nonmilitary males and females receiving directive supervision.

The researcher now has good reason to feel fairly confident of the findings, and might implement directive supervision across the district. The researcher also knows that if there is ever a choice, it is a safer best to use directive supervision with men and with ex-military men and women.

<b>RESEARCH QUESTION 3. DO DIFFERENT STYLES OF PRINCIPAL LEADERSHIP AFFECT TEACHER SATISFACTION?</b>
<b>Level 1 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction</b>
<p>Based on research he has read and observations he has made, a researcher has a hunch that there are different styles of school (principal) leadership, and that these styles are associated with different degrees of teachers' job satisfaction.</p> <p>This hunch is certainly not enough even to <b>suggest</b> that principals use a certain leadership style. Basic research is needed to see if there is anything TO the hunch.</p> <p>The researcher distributes <b>structured questionnaires</b> (a series of questions that all teachers are asked) to six schools (two elementary, middle, and high schools) in a district to see <b>if there is any association</b> (correlation) between teachers' evaluations of school leadership and teachers' job satisfaction).</p>
<b>Level 1 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction</b>
<b>Level 2 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction</b>
<p>Let's say that the <b>researcher finds from Level 1 research</b> that there is an association (correlation) between teachers' evaluations of school leadership and teachers' job satisfaction. For instance, teachers were most satisfied when principals had the following set of traits:</p> <ol style="list-style-type: none"> <li>Establishes a clear and feasible school mission that organizes all aspects of the school around proficient instruction.</li> <li>Is well-read in current scientific research. Shares this with staff.</li> <li>Skillfully evaluates and helps teachers to select effective curriculum materials.</li> <li>Skillfully models and provides useful suggestions for effective instruction methods when teachers need help.</li> <li>Provides useful resources for assessment, instruction, and remediation---workshops, materials.</li> <li>Listens and finds useful points in differing opinions, but makes and takes responsibility for final decisions about curriculum, instruction, supervision, etc.</li> </ol>



g. Visits all teachers in class almost every day, and praises teachers in front of class.

However, the sample was not large, and was not representative of all schools. Therefore, the researcher decides to test whether the findings “hold up” in other settings.

The researcher turns the findings into an hypothesis to test more rigorously.

“Teachers will be most satisfied when principals have a leadership style that involves the following features:....(described above).”

**Level 3 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction**

**c. LEVELS OF RESEARCH, WITH EXAMPLES**

**RESEARCH QUESTION 1. CAN WE IMPROVE THE ACHIEVEMENT OF DIVERSE LEARNERS?**

**Level 1 Research on Improving Achievement of Diverse Learners**

Two years ago, the Sunnyside County School District replaced largely ineffective elementary grades reading and math curricula with scientifically tested and effective curricula.

The district’s **official statistics now** show that at least 85% of White and/or affluent students read proficiently and pass the state math tests in grade 4---a big improvement. However, 50% of minority and poor students **STILL** read at the “below basic level” at the end of grade 4; and 60% fail the fourth grade state math test. These students will do poorly in all other subjects, and many will drop out of school.

In other words, although teachers now use curricula known to be effective, most diverse students are still failing.

The district knows research showing that diverse learners **CAN** achieve at high levels. Therefore, school failure can’t be explained away by poverty, family background, or culture.

Administrators want to find out why so many diverse learners are doing so poorly (what variables are **associated** with achievement vs. lack of achievement), and what variables can be **changed** that will **improve** education outcomes.

Does anything in early elementary grade classrooms **predict** a path of academic success vs. failure?

The researchers don't know for sure what to look for. So, they begin with data collection methods that provide **a lot of information of different kinds---field notes**. [In other words, they cast a broad net.] Also, it makes no sense at this point to study all schools. It's not wise to spend so much time and effort if you don't know what to look for.

The researchers observe in **two schools**: (1) One whose population is predominantly diverse learners.

(2) One that has a small population of diverse learners, but where there are at least five diverse learners in each classroom.

Twelve classes are studied in each school--two classes each in grades k-5.

The researchers want to see if they get any **clues** about what variables are associated with low vs. higher achievement in diverse learners.

Observers **take field notes** one day a week for one hour (varying the time of day) for a school year.

By studying students in the 6 grades (k-5) at once (over a year), the observers can see if there are changes in students **within** each grade from the start to the end of the year. They can also see if students are different from one grade level to the next. For example, does achievement go down in the higher grades?

Because the study is done for a year, it is a **longitudinal** (over time) **study**. Notice, researchers are **NOT** studying the **same group** of students (called a **panel**) as they move together from kindergarten to the end of grade 5. This would take six years. They don't HAVE six years to improve achievement! Instead, they study classes **AT** each grade level at the same time. **Each class is called a cohort**.

Observers are guided by past research that **SUGGESTS** the importance of certain variables. These are:

(1) Where different students sit.

(2) How often the teacher gives students opportunities to respond.

(3) How often the teacher gives praise and other forms of recognition and acceptance.

(4) Whether the teacher provides adequate wait time.

(5) Whether the teacher teaches and prompts students to use important and general pre-skills, such as listening for important information, remembering information and taking notes, vocabulary.

(6) How patient the teacher seems to be, as judged by facial expression and tone of voice.

(7) How the teacher responds to off-task or disruptive behavior.

The researchers analyze the field notes **within** each grade level. They look for changes in student achievement from the beginning to the end of each year.

They also compare findings **across** the 6 cohort grade levels, to see if, for example, there are changes as students move from kindergarten to grade 5.

The researchers find that **there are modest associations between teacher-student interaction and achievement in diverse learners**. For instance,

(1) At every grade level (k-5), the majority of diverse learners with **higher** achievement were in classes where teachers made sure they sat up front; gave lots of opportunities to respond; were patient; gave frequent praise; taught important pre-skills and responded to off-task and disruptive behavior in a matter-of-fact way.

(2) The majority of diverse learners who were falling behind in each grade level (becoming disengaged, making many errors, learning very little in reading, math, and other subjects), were in classes where the teachers did not teach in the above way.

(3) **At each grade level**, diverse learners in classes where teachers **did not** make sure they sat up front; **did not** give lots of opportunities to respond; were not patient; **did not** give frequent praise; **did not** teach important pre-skills; and **did not** respond to off-task and disruptive behavior in a matter-of-fact way, became LESS engaged and made less and less progress from the beginning to the end of the year.

(4) **Comparing the grade cohorts**, researchers found that in the higher grades, diverse learners started the year already less engaged and with less background knowledge needed for the new class, and began to fall behind right away.

The researchers are encouraged that they have discovered a possible association between specific features of instruction and student engagement and achievement. However, the sample was small. The findings may be a fluke. So, the researchers decide to move to Level 2 research. [See the section below.]

Level 2 Research on Improving Achievement of Diverse Learners
Level 3 Research on Improving Achievement of Diverse Learners
RESEARCH QUESTION 2. IS ONE FORM OF SUPERVISION (DIRECTIVE VS. NONDIRECTIVE) MORE EFFECTIVE THAN ANOTHER?
Level 1 Research on the Possible Effects of Directive vs. Nondirective Teacher Supervision.
<p>For instance, a supervisor has a hunch (from earlier observations or from several research articles that she read), that teachers may benefit more from directive supervision, with explicit instruction from the supervisor.</p> <p><b>She does pilot research to see if there is anything to this hunch.</b> She provides highly directive supervision to five new teachers, and <b>nondirective</b> supervision (teachers evaluate themselves and suggest ways to improve their teaching) to five other teachers---to see if there is <b>any consistent different</b> in these teachers' improvement.</p> <p>If the researcher <b>DOES</b> find the expected consistent differences, she next uses Level 2 research to test more <b>rigorously</b> (i.e., using design elements cited in the left-hand column) her hypothesis (not merely a hunch) on the effects of directive supervision</p>
Level 2 Research on the Possible Effects of Directive vs. Nondirective Teacher Supervision.

Let's say that the **researcher finds from Level 1 research** that teachers who received more directive supervision improved instruction more than teachers who received **nondirective** supervision.

The researcher turns this finding into an **hypothesis** to test more rigorously.

"Teachers who receive more directive supervision will improve their instruction and will be more satisfied with supervision than teachers who receive nondirective supervision." [Notice that several outcome variables are measured, to get a bigger picture of the possible effects of the

The researcher:

- a. Develops a **conceptual definition** of "proficient instruction." Then she transforms this definition into an **operational definition**---specific features of proficient instruction; e.g., using clear words, using a proper set of examples, correcting errors.
- b. Develops a **conceptual definition** of "teacher satisfaction with supervision." Then she transforms this definition into an **operational definition**---specific things that teachers might do and say, that represent different degrees of satisfaction with supervision.
- c. Develops **instruments** (See **DATA** in the left-hand column.) for **rating** teachers' "proficient instruction" and satisfaction with supervision."
- d. Tests the **reliability** of the instruments to see if different observers score the same teachers' proficiency and satisfaction the same way.
- e. Develops a **protocol** that tells exactly how to deliver directive vs. nondirective supervision
- f. Trains some supervisors in two suburban schools to supervise in a directive way and others to supervise in a nondirective way.
- g. Randomly assigns **20 teachers** to the two supervision groups.
- h. Pre-tests (rates) teachers' proficiency at instruction.
- i. Implements the two kinds of supervision.
- j. Post-tests (rates) teachers'

k. Uses **structured questionnaires** or **interviews** to find out how satisfied teachers were with their kind of supervision.

l. Analyzes the data to see if the data support the hypothesis; that is, do the data say that teachers who received directive supervision improved more (and were more satisfied with supervision) than teachers who received nondirective supervision?

For example,

What is the average amount of change in skill for each feature of proficient instruction in teachers who received directive vs. nondirective supervision?

What percentage of teachers who received directive supervision vs. nondirective supervision rated themselves as “highly satisfied”?

### Level 3 Research on the Possible Effects of Directive vs. Nondirective Teacher Supervision.

Let’s say that the researcher finds from Level 2 research that the data support the hypothesis.

“Sure enough, teachers who received more directive supervision DID improve their instruction and WERE more satisfied than teachers who received nondirective supervision.”

Great!

But is one study of 20 teachers in two suburban schools enough to be **confident** that there would be the **same results with different samples** of teachers (older, male, different grades); in different schools (urban, rural) and with different supervisors? No.

Is it ethical for the researcher to say that all schools ought to supervise in a directive way? No.

The researcher can’t be sure that the findings in the first study WEREN’T a chance fluke (that is, wouldn’t happen again if the study could be repeated exactly).

And the researcher can’t be sure that the findings apply anywhere else.

Therefore, the researcher conducts **larger-scale evaluation research**. For instance, the researcher:

a. Does the study again with samples that are similar to the first ones, to see if the results of the first study were a fluke .

[Let's say there are the same results. Probably not a fluke.]

b. Does the study across the whole district. There are 40 elementary schools.

The schools are **randomly assigned** (by flip of a coin) to the two supervision groups.

A **random sample** of 10 teachers is selected from each school.

Observers are **trained** to use the instruments for measuring proficient instruction and satisfaction with supervision. The **reliability** of their scorings is checked.

Supervisors in the schools are trained to use the protocols for providing directive vs.nondirective supervision.

Teachers receive **pretests** of their teaching skills; then teachers receive one of the two kinds of supervision; and then teachers receive post-tests of their teaching skill.

Teachers are given structured questionnaires or interviews to find out how satisfied teachers were with their kind of supervision.

The researcher analyzes the data. For example,

What is the average amount of change in skill for each feature of proficient instruction in teachers who received directive vs. nondirective supervision?

What percentage of teachers who received directive supervision vs. nondirective supervision rated themselves as “highly satisfied”?

The researcher finds that, again, directive supervision yields significantly more improvement in teaching skill and significantly higher ratings of satisfaction.

However, the researcher now **disaggregates** the data.

She divides the data on all teachers into groups whose characteristics may be relevant: male/female; veteran/new teachers; ex-military/non military.

Does the teacher's sex, time as a teacher, or military experience influence the effectiveness of the two forms of supervision?

The researcher finds that in general, directive supervision STILL works best---whether male or female, new or veteran, military experience or not.

However, male teachers and ex-military (males and females) receiving directive supervision **improve more** than female teachers and nonmilitary males and females receiving directive supervision.

The researcher now has good reason to feel fairly confident of the findings, and might implement directive supervision across the district. The researcher also knows that if there is ever a choice, it is a safer best to use directive supervision with men and with ex-military men and women.

### **RESEARCH QUESTION 3. DO DIFFERENT STYLES OF PRINCIPAL LEADERSHIP AFFECT TEACHER SATISFACTION?**

#### **Level 1 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction**

Based on research he has read and observations he has made, a researcher has a hunch that there are different styles of school (principal) leadership, and that these styles are associated with different degrees of teachers' job satisfaction.

This hunch is certainly not enough even to **suggest** that principals use a certain leadership style. Basic research is needed to see if there is anything TO the hunch.

The researcher distributes **structured questionnaires** (a series of questions that all teachers are asked) to six schools (two elementary, middle, and high schools) in a district to see if **there is any association** (correlation) between teachers' evaluations of school leadership and teachers' job satisfaction).

#### **Level 1 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction**

#### **Level 2 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction**



Let's say that the **researcher finds from Level 1 research** that there is an association (correlation) between teachers' evaluations of school leadership and teachers' job satisfaction. For instance, teachers were most satisfied when principals had the following set of traits:

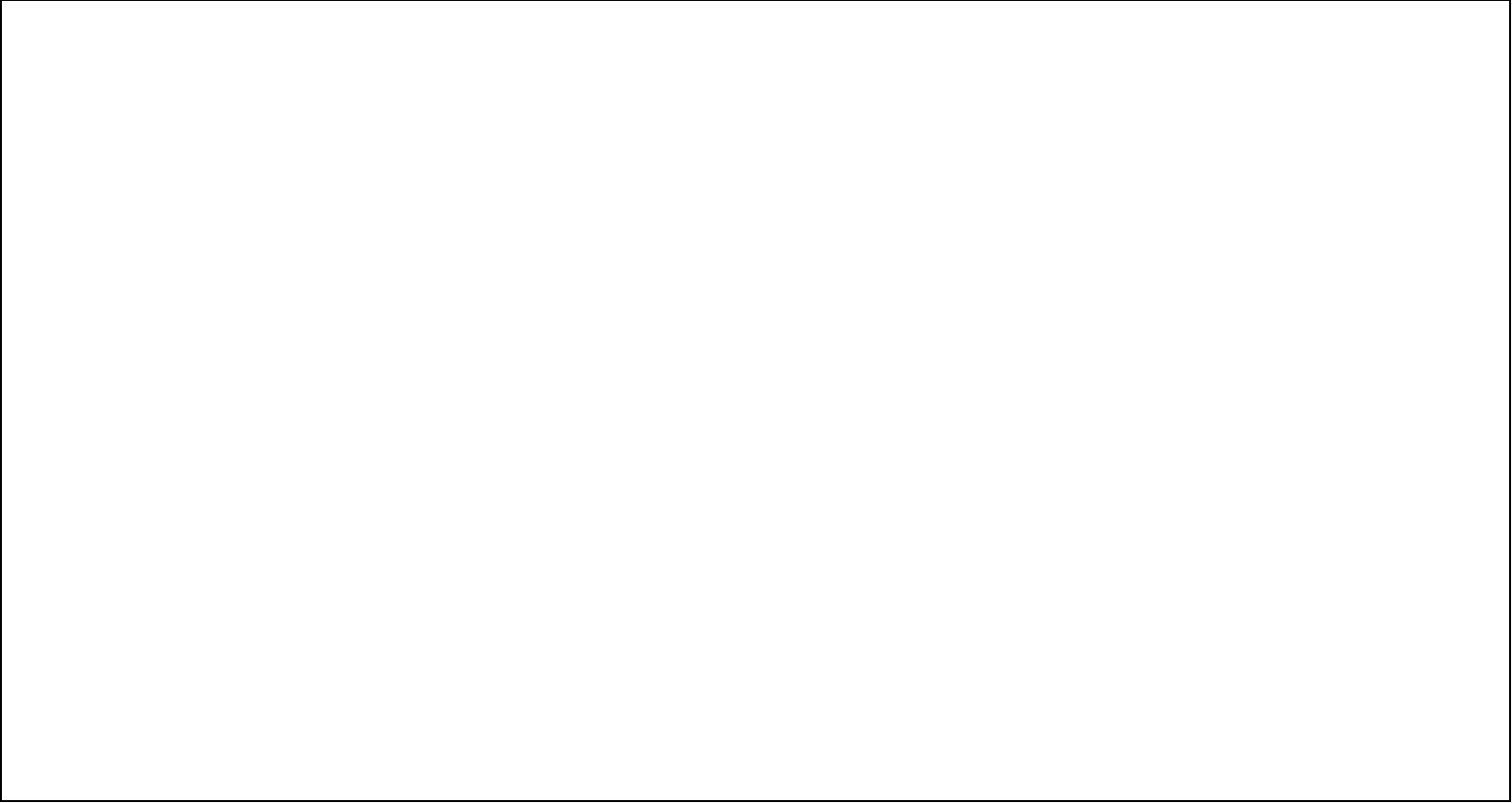
- a. Establishes a clear and feasible school mission that organizes all aspects of the school around proficient instruction.
- b. Is well-read in current scientific research. Shares this with staff.
- c. Skillfully evaluates and helps teachers to select effective curriculum materials.
- d. Skillfully models and provides useful suggestions for effective instruction methods when teachers need help.
- e. Provides useful resources for assessment, instruction, and remediation---workshops, materials.
- f. Listens and finds useful points in differing opinions, but makes and takes responsibility for final decisions about curriculum, instruction, supervision, etc.
- g. Visits all teachers in class almost every day, and praises teachers in front of class.

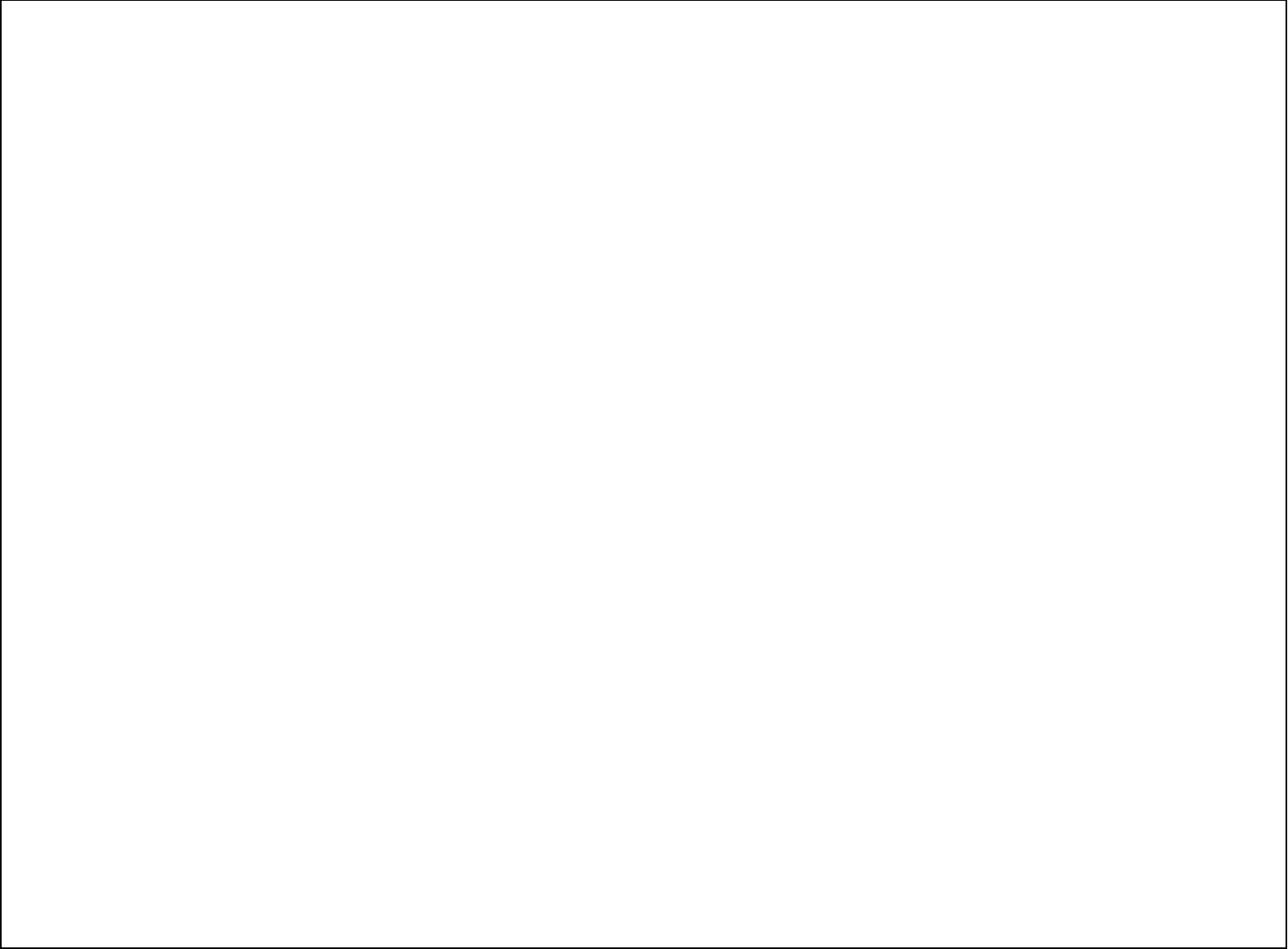
However, the sample was not large, and was not representative of all schools. Therefore, the researcher decides to test whether the findings "hold up" in other settings.

The researcher turns the findings into an hypothesis to test more rigorously.

"Teachers will be most satisfied when principals have a leadership style that involves the following features:....(described above)."

### **Level 3 Research on the Possible Effects of Different Styles of Principal Leadership Affect Teacher Satisfaction**



LEVEL 1 →	LEVEL 2 →	LEVEL 3
<b>Improving Instruction and Achievement for Diverse Learners</b>	<b>Improving Instruction and Achievement for Diverse Learners</b>	<b>Improving Instruction and Achievement for Diverse Learners</b>
<p>Two years ago, the Sunnyside County School District replaced largely ineffective elementary grades reading and math curricula with scientifically tested and effective curricula.</p> <p>The district's <b>official statistics</b> now show that at least 85% of White and/or affluent students read proficiently and pass the state math tests in grade 4---a big improvement. However, 50% of minority and poor students <b>STILL</b> read at the "below basic level" at the end of grade 4; and 60% fail the fourth grade state math test. These students will do poorly in all other subjects, and many will drop out of school.</p> <p>In other words, although teachers now use curricula known to be effective, most diverse students are still failing.</p> <p>The district knows research showing that diverse learners <b>CAN</b> achieve at high levels. Therefore, school failure can't be explained away by poverty, family background, or culture.</p> <p>Administrators want to find out why so many diverse learners are doing so poorly (what variables are <b>associated</b> with achievement vs. lack of achievement), and what variables can be <b>changed</b> that</p>	<p>The researchers transform their finding into an hypothesis</p>	

<p>will <b>improve</b> education outcomes.</p> <p>Does anything in early elementary grade classrooms <b>predict</b> a path of academic success vs. failure?</p> <p>The researchers don't know for sure what to look for. So, they begin with data collection methods that provide a <b>lot of information of different kinds---field notes</b>. [In other words, they cast a broad net.] Also, it makes no sense at this point to study all schools. It's not wise to spend so much time and effort if you don't know what to look for.</p> <p>The researchers observe in <b>two schools</b>:</p> <p>(1) One whose population is predominantly diverse learners.</p> <p>(2) One that has a small population of diverse learners, but where there are at least five diverse learners in each classroom.</p> <p>Twelve classes are studied in each school--two classes each in grades k-5.</p> <p>The researchers want to see if they get any <b>clues</b> about what variables are associated with low vs. higher achievement in diverse learners.</p> <p>Observers <b>take field notes</b> one day a week for one hour (varying the time of day) for a school year.</p>		
---	--	--

<p>By studying students in the 6 grades (k-5) at once (over a year), the observers can see if there are changes in students <b>within</b> each grade from the start to the end of the year. They can also see if students are different from one grade level to the next. For example, does achievement go down in the higher grades?</p> <p>Because the study is done for a year, it is a <b>longitudinal</b> (over time) <b>study</b>. Notice, researchers are <b>NOT</b> studying the <b>same group</b> of students (called a <b>panel</b>) as they move together from kindergarten to the end of grade 5. This would take six years. They don't HAVE six years to improve achievement! Instead, they study classes <b>AT</b> each grade level at the same time. <b>Each class is called a cohort.</b></p> <p>Observers are guided by past research that <b>SUGGESTS</b> the importance of certain variables. These are:</p> <p>(1) Where different students sit.</p> <p>(2) How often the teacher gives students opportunities to respond.</p> <p>(3) How often the teacher gives praise and other forms of recognition and acceptance.</p> <p>(4) Whether the teacher provides adequate wait time.</p>		
---	--	--

<p>(5) Whether the teacher teaches and prompts students to use important and general pre-skills, such as listening for important information, remembering information and taking notes, vocabulary.</p> <p>(6) How patient the teacher seems to be, as judged by facial expression and tone of voice.</p> <p>(7) How the teacher responds to off-task or disruptive behavior.</p> <p>The researchers analyze the field notes <b>within</b> each grade level. They look for changes in student achievement from the beginning to the end of each year.</p> <p>They also compare findings <b>across</b> the 6 cohort grade levels, to see if, for example, there are changes as students move from kindergarten to grade 5.</p> <p>The researchers find that <b>there are modest associations between teacher-student interaction and achievement in diverse learners</b>. For instance,  (1) At every grade level (k-5), the majority of diverse learners with <b>higher</b> achievement were in classes where teachers made sure they sat up front; gave lots of opportunities to respond; were patient; gave frequent praise; taught important pre-skills and responded to off-task and disruptive behavior in a matter-of-fact way.</p>		
---	--	--

<p>(2) The majority of diverse learners who were falling behind in each grade level (becoming disengaged, making many errors, learning very little in reading, math, and other subjects), were in classes where the teachers did not teach in the above way.</p> <p>(3) <b>At each grade level</b>, diverse learners in classes where teachers <b>did not</b> make sure they sat up front; <b>did not</b> give lots of opportunities to respond; were not patient; <b>did not</b> give frequent praise; <b>did not</b> teach important pre-skills; and <b>did not</b> respond to off-task and disruptive behavior in a matter-of-fact way, became LESS engaged and made less and less progress from the beginning to the end of the year.</p> <p>(4) <b>Comparing the grade cohorts</b>, researchers found that in the higher grades, diverse learners started the year already less engaged and with less background knowledge needed for the new class, and began to fall behind right away.</p> <p>The researchers are encouraged that they have discovered a possible association between specific features of instruction and student engagement and achievement. However, the sample was small. The findings may be a fluke. So, the researchers decide to move to Level 2 research. [See the column on the</p>		
---	--	--



right.]		
<b>LEVEL 1 →</b>	<b>LEVEL 2 →</b>	<b>LEVEL 3</b>
<b>The effects of directive and nondirective teacher supervision.</b>	<b>The effects of directive and nondirective teacher supervision.</b>	<b>The effects of directive and nondirective teacher supervision.</b>
<p>May use <b>experimental designs</b>, but (since this is <b>basic research</b> or a <b>pilot study</b>) the experiment may not have equivalent control and comparison groups and standardized measurement.</p> <p>For instance, a supervisor has a hunch (from earlier observations or from several research articles that she read), that teachers may benefit more from directive supervision, with explicit instruction from the supervisor.</p> <p><b>She does pilot research to see if there is anything to this hunch.</b> She provides highly directive supervision to five new teachers, and <b>nondirective</b> supervision (teachers evaluate themselves and suggest ways to improve their teaching) to five other teachers---to see if there is <b>any consistent different</b> in these teachers' improvement.</p> <p>If the researcher DOES find the expected consistent differences, she next uses Level 2 research to test more <b>rigorously</b> (i.e., using design elements cited in the</p>	<p>Let's say that the <b>researcher finds from Level 1 research</b> that teachers who received more directive supervision improved instruction more than teachers who received <b>nondirective</b> supervision.</p> <p>The researcher turns this finding into an <b>hypothesis</b> to test more rigorously.</p> <p>"Teachers who receive more directive supervision will improve their instruction and will be more satisfied with supervision than teachers who receive nondirective supervision." [Notice that several outcome variables are measured, to get a bigger picture of the possible effects of the</p> <p>The researcher: a. Develops a <b>conceptual definition</b> of "proficient instruction." Then she transforms this definition into an <b>operational definition</b>---specific features of proficient instruction; e.g., using clear words, using a proper set of examples, correcting errors.</p>	<p>Let's say that the researcher finds from Level 2 research that the data support the hypothesis.</p> <p>"Sure enough, teachers who received more directive supervision DID improve their instruction and WERE more satisfied than teachers who received nondirective supervision."</p> <p>Great!</p> <p>But is one study of 20 teachers in two suburban schools enough to be <b>confident</b> that there would be the <b>same results with different samples</b> of teachers (older, male, different grades); in different schools (urban, rural) and with different supervisors? No.</p> <p>Is it ethical for the researcher to say that all schools ought to supervise in a directive way? No.</p> <p>The researcher can't be sure that the findings in the first study WEREN'T a</p>

<p>left-hand column) her hypothesis (not merely a hunch) on the effects of directive supervision</p>	<p>b. Develops a <b>conceptual definition</b> of “teacher satisfaction with supervision.” Then she transforms this definition into an <b>operational definition</b>---specific things that teachers might do and say, that represent different degrees of satisfaction with supervision.</p> <p>c. Develops <b>instruments</b> (See <b>DATA</b> in the left-hand column.) for <b>rating</b> teachers’ “proficient instruction” and satisfaction with supervision.”</p> <p>d. Tests the <b>reliability</b> of the instruments to see if different observers score the same teachers’ proficiency and satisfaction the same way.</p> <p>e. Develops a <b>protocol</b> that tells exactly how to deliver directive vs. nondirective supervision</p> <p>f. Trains some supervisors in two suburban schools to supervise in a directive way and others to supervise in a nondirective way.</p> <p>g. Randomly assigns <b>20 teachers</b> to the two supervision groups.</p> <p>h. Pre-tests (rates) teachers’ proficiency at instruction.</p> <p>i. Implements the two kinds of supervision.</p> <p>j. Post-tests (rates) teachers’</p>	<p>chance fluke (that is, wouldn’t happen again if the study could be repeated exactly).</p> <p>And the researcher can’t be sure that the findings apply anywhere else.</p> <p>Therefore, the researcher conducts <b>larger-scale evaluation research</b>. For instance, the researcher:</p> <p>a. Does the study again with samples that are similar to the first ones, to see if the results of the first study were a fluke .</p> <p>[Let’s say there are the same results. Probably not a fluke.]</p> <p>b. Does the study across the whole district. There are 40 elementary schools.</p> <p>The schools are <b>randomly assigned</b> (by flip of a coin) to the two supervision groups.</p> <p>A <b>random sample</b> of 10 teachers is selected from each school.</p> <p>Observers are <b>trained</b> to use the instruments for measuring proficient instruction and satisfaction with supervision. The <b>reliability</b> of their scorings is checked.</p> <p>Supervisors in the schools are trained</p>
--	--	--

	<p>k. Uses <b>structured questionnaires</b> or <b>interviews</b> to find out how satisfied teachers were with their kind of supervision.</p> <p>l. Analyzes the data to see if the data support the hypothesis; that is, do the data say that teachers who received directive supervision improved more (and were more satisfied with supervision) than teachers who received nondirective supervision?</p> <p>For example,</p> <p>What is the average amount of change in skill for each feature of proficient instruction in teachers who received directive vs. nondirective supervision?</p> <p>What percentage of teachers who received directive supervision vs. nondirective supervision rated themselves as “highly satisfied”?</p>	<p>to use the protocols for providing directive vs. nondirective supervision.</p> <p>Teachers receive <b>pretests</b> of their teaching skills; then teachers receive one of the two kinds of supervision; and then teachers receive post-tests of their teaching skill.</p> <p>Teachers are given structured questionnaires or interviews to find out how satisfied teachers were with their kind of supervision.</p> <p>The researcher analyzes the data. For example,</p> <p>What is the average amount of change in skill for each feature of proficient instruction in teachers who received directive vs. nondirective supervision?</p> <p>What percentage of teachers who received directive supervision vs. nondirective supervision rated themselves as “highly satisfied”?</p> <p>The researcher finds that, again, directive supervision yields significantly more improvement in teaching skill and significantly higher ratings of satisfaction.</p> <p>However, the researcher now <b>disaggregates</b> the data.</p>
--	---	---

		<p>She divides the data on all teachers into groups whose characteristics may be relevant: male/female; veteran/new teachers; ex-military/non military.</p> <p>Does the teacher's sex, time as a teacher, or military experience influence the effectiveness of the two forms of supervision?</p> <p>The researcher finds that in general, directive supervision STILL works best--whether male or female, new or veteran, military experience or not.</p> <p>However, male teachers and ex-military (males and females) receiving directive supervision <b>improve more</b> than female teachers and nonmilitary males and females receiving directive supervision.</p> <p>The researcher now has good reason to feel fairly confident of the findings, and might implement directive supervision across the district. The researcher also knows that if there is ever a choice, it is a safer best to use directive supervision with men and with ex-military men and women.</p>
LEVEL 1 →	LEVEL 2 →	LEVEL 3
The effects of principal leadership style on teacher satisfaction	The effects of principal leadership style on teacher satisfaction	The effects of principal leadership style on teacher satisfaction

<p>Based on research he has read and observations he has made, a researcher has a hunch that there are different styles of school (principal) leadership, and that these styles are associated with different degrees of teachers' job satisfaction.</p> <p>This hunch is certainly not enough even to <b>suggest</b> that principals use a certain leadership style. Basic research is needed to see if there is anything TO the hunch.</p> <p>The researcher distributes <b>structured questionnaires</b> (a series of questions that all teachers are asked) to six schools (two elementary, middle, and high schools) in a district to see <b>if there is any association</b> (correlation) between teachers' evaluations of school leadership and teachers' job satisfaction).</p>	<p>Let's say that the <b>researcher finds from Level 1 research</b> that there is an association (correlation) between teachers' evaluations of school leadership and teachers' job satisfaction. For instance, teachers were most satisfied when principals had the following set of traits:</p> <ul style="list-style-type: none"> <li>a. Establishes a clear and feasible school mission that organizes all aspects of the school around proficient instruction.</li> <li>b. Is well-read in current scientific research. Shares this with staff.</li> <li>c. Skillfully evaluates and helps teachers to select effective curriculum materials.</li> <li>d. Skillfully models and provides useful suggestions for effective instruction methods when teachers need help.</li> <li>e. Provides useful resources for assessment, instruction, and remediation---workshops, materials.</li> <li>f. Listens and finds useful points in differing opinions, but makes and takes responsibility for final decisions about curriculum, instruction, supervision, etc.</li> <li>g. Visits all teachers in class almost every day, and praises teachers in front of</li> </ul>	

	<p>class.</p> <p>However, the sample was not large, and was not representative of all schools. Therefore, the researcher decides to test whether the findings “hold up” in other settings.</p> <p>The researcher turns the findings into an hypothesis to test more rigorously.</p> <p>“Teachers will be most satisfied when principals have a leadership style that involves the following features:....(described above).”</p>	

- Anderson, J.R., Reder, L.M., & Simon, H.A. (1998). Applications and Misapplications of Cognitive Psychology to Mathematics Education. Department of Psychology. Carnegie Mellon University. Pittsburgh, PA 15213.
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19, 163-197.
- Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Witrock (Ed.), *Third handbook of research on teaching* (pp. 328-375). New York: McMillan.
- Carnine, D. W. (1976). *Correction effects on academic performance during small group instruction*. Unpublished manuscript. Eugene, OR: University of Oregon Follow Through Project.
- Dixon, R. (1989). Sequences of Instruction. University of Oregon.
- Dixon, R.C. , & Carnine, D. (1993). Using scaffolding to teach writing. *Educational Leadership*, 51 (3), 100-101.

- Dougherty, K.M., & Johnston, J.M. (1996). Overlearning, fluency, and automaticity. *The Behavior Analyst*, 19, 289-292.
- Ehri, L.C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading* (pp. 3-40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ehri, L. (2002). Phases of acquisition in learning to read words and implications for teaching. In R. Stainthorp and P. Tomlinson (Eds.) *Learning and teaching reading*. London: British Journal of Educational Psychology Monograph Series II.
- Ellis, E.S., & Worthington, L.A. (1994). *Research synthesis on effective teaching principles and the design of quality tools for educators*. University of Oregon: National Center to Improve the Tools of Educators.
- Engelmann, S. (1999). Student program alignment and teaching to mastery. National Direct Instruction Conference. Eugene, Oregon.
- Englert, C.S., Raphael, T.E., Anderson, L.M., Anthony, H.M., & Stevens, D.D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal*, 2, 337-372.
- Greenwood, C.R., Delquadri, J., & Hall, R.V. (1984). Opportunity to respond and student academic performance. In W.L. Heward, T.E. Heron, J. Trap-Porter, & D.S. Hill (Eds.), *Focus on behavior analysis in education*. Columbus, OH: Merrill.
- Grossen, B.J., Carnine, D.W., Romance, N.R., & Vitale, M.R. (1998). Effective strategies for teaching science, in E.J. Kameenui & D.W. Carnine (Eds.), *Effective teaching strategies that accomodate diverse learners*, pp. 113-137. Columbus, OH. Merrill.
- Grossen, B. (1996). How should we group to achieve excellence with equity? Eugene, OR: University of Oregon. On-line at <http://darkwing.uoregon.edu/~adiep/grp.htm>
- Gunn, B.K., Simmons D.C., & Kameenui, E.J. (1995b). *Emergent literacy: Curricular and instructional implications for diverse learners*. Eugene, OR: University of Oregon, National Center to Improve the Tools of Educators. [On-Line]. Available: <http://idea.uoregon.edu/~ncite/documents/techrep/tech20.html>
- Gunter, P.L., Hummel, J.H., & Conroy, M.A. (1998). Increasing correct academic responding: An effective intervention strategy to decrease behavior problems. *Effective School Practices*, 17, 2, 55-62.
- Haring, N.G., White, O.R., & Liberty, K.A. (1978). *An investigation of phases of learning and facilitating instructional events for the severely handicapped. An annual progress report, 1977-78*. Bureau of Education of the Handicapped, Project No. 443CH70564. Seattle: University of Washington, College of Education.

- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437-447.
- Kame'enui, E.J., & Simmons, D.C. (1990). *Designing instructional strategies: The prevention of academic learning problems*. Columbus, OH: Merrill.
- Kame'enui, E.J., & Simmons, D.C. (1999). *Toward successful inclusion of students with disabilities: The architecture of instruction*. Reston, VA: Council for Exceptional Children.
- Kame'enui et al., (2002). *Analysis of reading assessment instrument K-3 Summary of assessment committee decisions Grade level by type of measure*.
- Romance, N.R., & Vitale, M.R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education*, 23, (4), 373-404.
- Rosenshein, B., & Meister, C. (1992). The use of scaffolds for teaching higher-order cognitive strategies. *Educational Leadership*, 49 (7), 26-33.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (Third edition) (pp. 376-391). New York: McMillan..
- Rosenshine, B. (1986). Synthesis of research on explicit teaching. *Educational Leadership*, 43, 60-69.
- Stein, M., Kidder, D., Silbert, J., & Carnine, D. (2006). *Designing effective mathematics instruction*. (Third edition). Upper Saddle River, NJ: Pearson. Merrill.
- Stevenson, H.W. & Stigler, J.W. (1992). *The learning gap : why our schools are failing and what we can learn from Japanese and Chinese education*. New York : Summit Books.
- Walberg, H.J. (1990). Productive teaching and instruction: Assessing the knowledge base. *Phi Delta Kappan*, February, 470-478.