Research Vocabulary

Martin Kozloff

Achievement. The amount of learning with respect to an objective from earlier to later measurements. Generally measured by scores on tests.

Achievement gap. Differences in achievement between subgroups, such as ethnic groups (White, Asian, African American, Latino, Native American) and economic classes (wealthy, middle class, poor).

Aggregate data. Data for a sample/group as a whole. For example, the average percentage of correct answers on a test for the whole (aggregate) group might be 75%

Best fit line. The best fit line is a line that shows the trend, or the shape of the change, or the relationship between values of one set of variables and values of another set of variables.

Comparison group. Groups in an experiment that differ in some way (e.g., curriculum, whose effects or outcomes are being investigated. One kind of comparison group is a "treatment" group (e.g., the group that receives a new curriculum). Another kind of comparison group is a "control group" (that does not receive the new curriculum). If two or more curricula, for example, are being tested against one another, then the two groups are "alternative treatment" groups.

Control groups. A control group is a group that does not receive an intervention that is being tested. The performance of the control group is compared with the performance of an experimental group that does receive the intervention to be tested. The experimental and control groups should be as similar as possible, so that the only significant difference is that the experimental group received the intervention and the control group did not. If there are differences in how much the experimental group changed between a pre-test and the post-test, over how much the control group changed, then, all other things (variables) being equal, the intervention probably made the difference.

If the experimental and control groups are NOT virtually equal in other ways besides the intervention, then you cannot conclude that it was the intervention that made the difference in, for example, achievement. [See Comparison group.]

Definitions (of variables). [See "variables."] A definition is a statement that tells what a word (a name for a variable, or concept) means, or signifies, or points to. If a definition clearly tells what a variable means, then you can more easily think of how to measure the variable---measure the events that it points to. For example, if fluency (a concept and a variable) means performance that is both accurate and rapid, then to measure fluency you must measure how accurately and rapidly a person does something.

Words don't tell you what they mean. Human beings invent definitions. There are two kinds of definitions.

Conceptual definitions. Conceptual definitions are broad. They are like a search light that shines on a general area. A conceptual definition of fluency might be:

> Fluency is a feature of performance: accuracy and speed.

Here is a conceptual definition of representative democracy.

> Representative democracy is a form of political system in which citizens have the right to vote elect representatives who make important local and societal decisions.

Notice that the conceptual definition of fluency directs your attention to two aspects of performance (accuracy and speed) and NOT to other aspects of performance, such as how independently persons performs a task, or how easily persons generalize knowledge or the performance to new situations. Likewise, the definition of representative democracy directs your attention to political systems that have certain features, and away from societies that have other features, such as dictatorships---where there is no voting.

Operational definitions. Conceptual definitions are not precise enough. To create actual ways of measuring a concept or variable, you need definitions that say EXACTLY what you would see or hear. For instance, an operational definition of fluent reading in grade 1 might be:

> By the end of grade 1, the student reads grade level connected text at the rate of 60 correct words per minute.

Notice that this operational definition DOES include accuracy and speed. But it is more precise than the conceptual definition. It is so precise that you can think of exactly how to measure fluency: grade 1 level connected text; the child reads the text; the observer marks errors; the child reads for one minute, the observer counts the number of errors and subtracts this from the total number of words read.

When you evaluate research, ask:

a. Did the writer provide conceptual definitions? For example, if a writer says that "teachers were trained," what does that mean? Trained to do what? What skills?

b. Did the writer provide operational definitions? For example, did the writer state how teachers were trained, how their learning was measured, how successful and unsuccessful performance was defined and measured? If not, then maybe different teachers were trained differently, and with different results. In other words, without operational definitions, **the word "trained" means nothing.**

c. Were conceptual definitions derived from or consistent with scientific research? For example, reading might be defined as

   The process of constructing meaning from text.

   Is that ALL that reading is? Comprehension alone? Scientific research shows that reading ALSO includes knowledge of the sounds that are associated with letters (phonics); using knowledge of letter-sounds to sound out words (decoding); hearing the separate sounds in words (phonemic awareness), and vocabulary (knowing the definitions of words). So, the above conceptual definition is narrow. It does not include enough of what is meant by reading in the scientific community. Any curriculum materials, instructional methods, and assessments/measures of reading will be INVALID.

d. Were operational definitions derived from and consistent with the conceptual definition? And did they include what is relevant to the concept and exclude what is irrelevant to the concept? For example, what exactly do you see or hear when someone constructs meaning from a text? Do they ask certain questions? Do they read on to check their answers? If that is part of the conceptual definition, then that is what should be in the operational definition.

> Reading is the process of constructing meaning from text using cognitive routines; for example, the reader asks questions such as (who, why, what, when)……..; and then reads on to check his or her answers; the reader restates sentences to himself or herself; the reader connects events into sequences.

This operational definition is better. It identifies what readers actually do. It includes what is important---at least for ONE aspect of reading (comprehension). And you can observe this! But isn't it a good idea for the operational definition explicitly to EXCLUDE guessing? A student might use GUESSING to construct meaning. And the student might be good at guessing. If the operational definition doesn't exclude guessing, then a student who guesses (rather than uses a cognitive routine) will be a proficient reader. Is guessing what reading usually means?

Disaggregation of data. A sample or group always has members who differ in certain ways: male/female/; White/Minority. Aggregate data for the whole groups don't tell about differences or similarities between the subgroups. To disaggregate data is to analyze data on subgroups of the sample. For example, the sample of all students who took an achievement test could be disaggregated (divided) into subgroups such as White, African American, Asian, Latino, and Native American. Then you can compare and contrast scores among the subgroups.

Diverse Learners. Learners from subgroups (ethnic, social class, learning difficulties) that bring less background knowledge (e.g., vocabulary, reading skill, reasoning strategies) to school and who may have a more difficult time learning, organizing, retrieving, and applying knowledge. These learners therefore require assessments that precisely identify their learning needs, and progress monitoring that enables teachers to provide supplemental, remedial, or intensive instruction.

Empirical, Empiricism. Empiricism is the central concept in scientific research or scientific thinking. It means that claims are based on empirical data---that is, data that come from **observation, from direct seeing or hearing**. Empiricism is in contrast to claims that are based on **speculation** ("I'm pretty sure that this new curriculum---Flapdoodle Phonics---works."), **hearsay** ("I was told by three teachers and two passing strangers that Flapdoodle Phonics works."); the **prestige of gurus** ("Professor Hindquarters advocates Flapdoodle Phonics. In fact, he invented it."); and **preferences** ("We like the pictures in Flapdoodle Phonics. Also the kids discover how to read all by themselves!").

When research is empirical---data come from direct observations (e.g., an observer counts the number of words students read correctly per minute), or from numbers that accurately describe a situation (e.g., school statistics on how many students passed a standardized test in math), certain thins are possible, that make claims (based on the empirical data) more believable.

1. Other persons can observe the same thing. Therefore, data (information) can be checked for accuracy.
2. Information can be collected the same (empirical) way again and again, so that a research question ("How many of our students are proficient at math now that we introduced a new program?") can be answered again and again (year after year).

3. Hypotheses and beliefs ("If we use the Mastery Math program, our students will achieve more than they did with the older program.") can be tested. Therefore, teachers will have solid information that they can use to make decisions---for instance, to continue to use Mastery Math, or not.

Equivalent groups. In order to see if an intervention has an effect, or in order to identify what factors (variables) make a difference, the groups (e.g., classes) being compared must be equivalent (nearly the same) in everything else. For example, if you want to see if a new math program raises achievement, and you give the new program to one class and the older program to another class, the two classes have to be equivalent in OTHER variables that might affect achievement. Otherwise, how could you tell if it was the program or the other variables that made the difference?

There are two ways to try to make groups equivalent.

Matching. Matching is one way to try to make experimental and control groups equivalent. You select variables (factors) that may have an effect on the thing you are measuring (e.g., achievement), and you make sure that the groups are similar in these variables. For example, the two groups are the same on the percentage of boys and girls; high and low income; and ethnic composition.

Randomization, or random allocation. This is a second way to try to make experimental and control groups equivalent. If you have a "pool" of 50 students, you randomly assign them to the two groups. This means that all factors (ethnicity, social class, family support, background knowledge, age, sex) have an equal chance of being in either group.

Experimental research. A research strategy that usually involves comparison of two or more situations (classes, schools, districts) that are the same in every way possible, but that differ in the factor/curriculum/method whose effects or outcomes you are testing. If there are differences in the outcomes or effects between a situation (comparison group) that received, for example, one curriculum, and a second situation (**comparison group**---in this case, **control group**) that did not, then, logically, the one major difference---the curriculum---made the difference in the outcome. See **Equivalent groups.**

Experimental groups. An experimental group is the group that receives the "intervention" (for example, new curriculum materials) whose effects are being assessed or tested.

Extraneous variables. Extraneous variables are variables that are not part of an intervention (e.g., a change in curriculum or instructional methods) whose effects are being tested. Extraneous variables may "interact with" independent (intervention) variables to produce an effect, or extraneous variables may produce an effect by themselves. Therefore, change (or lack of change) in dependent variables (e.g., reading achievement) may be entirely or partly the result of extraneous variables, such as maturation; other things happening outside of school (e.g., siblings teach some students to read); measurement error (students appear to read better because observers at the outcome assessment failed to count many errors); bias in selection (e.g., if the experimental group has many bright students and the control group doesn't, that different---and not the curriculum---may account for differences in achievement).

Hypothesis. A hypothesis is a statement of belief that can be tested. There are two kinds of hypotheses. The ***research hypothesis*** is what you believe to be the case; you collect data to see **if** the data support the research hypothesis. For example, you believe that adapting instruction to fit students' learning style is important. Your research hypothesis might be: "Students who receive math instruction that is consistent with their learning styles (experimental group) will make more gains during the year on math tests than students who do NOT receive math instruction that is consistent with their learning styles (control group)." You then assign students to the two groups (experimental and control group); give a pre-test of their math knowledge; give one group the adapted instruction and the other the usual instruction; give a post-test of their math knowledge; and determine if any differences are as predicted by your hypothesis. If so, the hypothesis is SUPPORTED. It is not PROVED to be TRUE, because OTHER things (errors of measurement, teacher behavior from one group to the other) might have raised the scores of the experimental group and held down the scores of the control group.

The other kind of hypothesis is the ***null hypothesis***. This is basically a statement of the opposite of the research hypothesis. For example, the null hypothesis might be "Students who receive math instruction that is consistent with their learning styles (experimental group) will make NO more gains during the year on math tests than students who do NOT receive math instruction that is consistent with their learning styles (control group)." You conduct the research as describe above. And if the findings are that students in the experimental group made more gains, then your null hypothesis is FALSE. This does not mean that the research hypothesis is true. It only means that IT is NOT false.

The null hypothesis is a way that researchers keep themselves honest. It is easy to FIND data that will support what you believe (your research hypothesis). The NULL hypothesis challenges the researcher to collect exactly the kind of data that SUPPORT the null hypothesis---that adapting instruction to learning styles makes NO difference.

Independent research. Research that is conducted by persons or groups who do not have a stake in the outcomes of the research. For example, it is not independent research is a publisher evaluates his own materials. There may be at least subtle bias in such research.

Levels of measurement. There are four levels of measurement: nominal, ordinal, interval, and ratio.

- **Nominal level.** Nominal measurement consists of naming or putting the things measured into categories. For example, you could categorize students into two groups: student who receive free and reduced lunch and students who do not receive free and reduced lunch. This nominal (name) measurement indicates difference in family income, but it is not precise information.
- **Ordinal level.** Ordinal measurement consists of placing the things measured into **ranks**. For example, teachers might observe students reading and then place each student in one of three ranks: Proficient/advanced, Basic, and Below basic. This ranking indicates differences in proficiency but, as with nominal measurement, it does not give precise information (such as how many correct words students read per minute). Also the differences between the ranks are not necessarily equal. That is, the difference in proficiency between Below basic and Basic, and between Basic and Proficient/advanced may not be equal. The difference in proficiency between Basic and Proficient/advanced may be far greater than the difference in proficiency between Below basic and Basic. Ordinal level measurement is sometimes provided by rating scales that ask persons to answer questions such as:

  How often would you say that you correct student errors?

  (1) Almost every time.

  (2) Most of the time.

  (3) Occasionally.

  (4) Rarely.
- **Interval level.** Interval level measurement is the kind of information provided by thermometers. There are a series of intervals (e.g., degrees) that are equal, and there is no true zero (there is no such thing as zero temperature). Interval level measurement is often provided by rating scales that ask persons to answer questions such as:

Place an X in the spot that best represents how teacher-friendly your new math materials are.

|____|____|____|____|

1 2 3 4

Less friendly More friendly

- **Ratio level.** Ratio level measurement is the most precise. It provides information of the number of times (e.g., number of questions answered correctly), or the rate (e.g., number of words read correctly per minute), or percentage of times (e.g., the percentage of errors teachers correct) that something happens. Ratio level information is usually provided through direct observation or through tests that enable the observer to instances of identified variables (e.g., correct answers).

Levels of research. There are three levels of research. There are also "research" claims that really are not ANY kind of research.

- **Nonresearch claims.** This is writing (e.g., articles) that merely asserts opinions, or beliefs, or "Most educators know that…," or "Piaget argued that…," or "According to constructivist philosophy…" There is little or no experimental test of the claims. Readers may be swayed merely because the writing uses emotionally charged and appealing language (holistic, seamless, natural, deep, everyone believes, child centered). Sometimes, the claims are called "theory," but they really are not theory. They are merely unsupported sentences about the writers' preferences for how children are taught. A true theory is a set of statements that are connected logically and that form a comprehensive explanation.

- **Level 1--Basic" research.** This research is field observations (e.g., observing peer reading exercises in class) or it involves some quantitative data (e.g., how many words each peer in the exercises reads correctly per minute when it is his or her turn). The research may be guided by an hypothesis of what the researcher thinks is the case (e.g., peer reading exercises increase reading fluency). The research identifies what APPEAR to be correlations. Or it shows that there are NO correlations. The research may provide a SOMEWHAT reasonable explanation (partial theory) for what is found.

- **Level 2--Test of the theory in real classrooms.** This research is more rigorous than level 1 research.

  a. Hypotheses are stated clearly.

  b. Variables in the hypotheses are clearly defined (e.g., exactly what goes on in the peer reading exercises, exactly what reading fluency means).

  c. Measures, and methods for making the measurements, are developed and tested to see if they are valid---measure what they are supposed to measure (See Validity). For example, reading experts are consulted on the definitions of fluency and the measures; e.g., each child reads a passage that is 100% decodable (the child knows how to read every word). Each child takes a turn reading. The other child, reading along, marks each error and checks how many minutes the reading took.
  In addition, the measures are checked for **reliability**. That is, if two observers measure the same child's fluency during an exercise, will the observers arrive at about the same score?

  d. Experimental and control groups are formed, and these groups are created by matching or by random allocation to try to ensure that the children are similar on variables that could influence reading fluency. The experimental group consists of students who do the peer reading exercises. The control group might be students who read by themselves and are given strategies for increasing fluency. [See Experimental group, Control group, and Matching.]

e. Fluency is measured at the beginning of the experimental TEST of the hypothesis, during each lesson, and at the end of the series, to see if there is any TREND in each group [See Trend.] and to see if (as hypothesized) the experimental group gains more in fluency than the control group.

f. Conclusions are drawn about whether the research hypothesis was supported and whether the null hypothesis (peer readers make no more gains than independent readers) can be rejected.

- **Level 3--Program Evaluation on a school- or district-wide basis.** The same rigorous research is done as in level 2. This research answers the question,

  > "Will we find the same thing (e.g., students who work on fluency in peer reading exercises DO make significantly higher gains---between pre-test and post-test---than students who work on fluency independently) when we do this at the level of a whole school or district?"

  In other words, level 3 research is checking the reliability (repeatability) of the results in different environments (e.g., with different children, and teachers, and different degrees of teacher support). It is one thing for a teacher to do the peer reading "protocol" (way of doing it) when she is in an experiment and is receiving special assistance to do it right. Bit what happens when peer reading exercises are just one part of the school activities? Will teachers use the protocol faithfully then? Level 3 research is what must be done BEFORE writers claim that an innovation works and should be used; and before teachers USE any new method. Would anyone use a drug that had only been tested/tired with 20 persons?

Longitudinal research. Longitudinal research is research done over a fairly long period of time. Research that is NOT longitudinal may show that a method is effective. However, you won't know if it is effective for very long.

Measure. A measure is simply information on the value of a variable. If reading proficiency is the variable, what is the measure of reading proficiency? That is, there is more or less of **<u>what</u>**? There can be many measures of a variable, because variables (such as reading proficiency) include a lot of things. For example, how many words (out of 20) does a child segment correctly ("What are the sounds in sun?")? How many letter-sound relationships (out of 40) does a child get right? [Teacher points to letters and says, "What sound?"] How many words (out of 100) does a child read correctly? How many words does a child read correctly in one minute? How many vocabulary words out of 100 does a child define correctly. How many questions (out of 20) about what a text says does a child answer correctly. These are all measures of reading proficiency.

NAEP. National Assessment of Educational Progress.

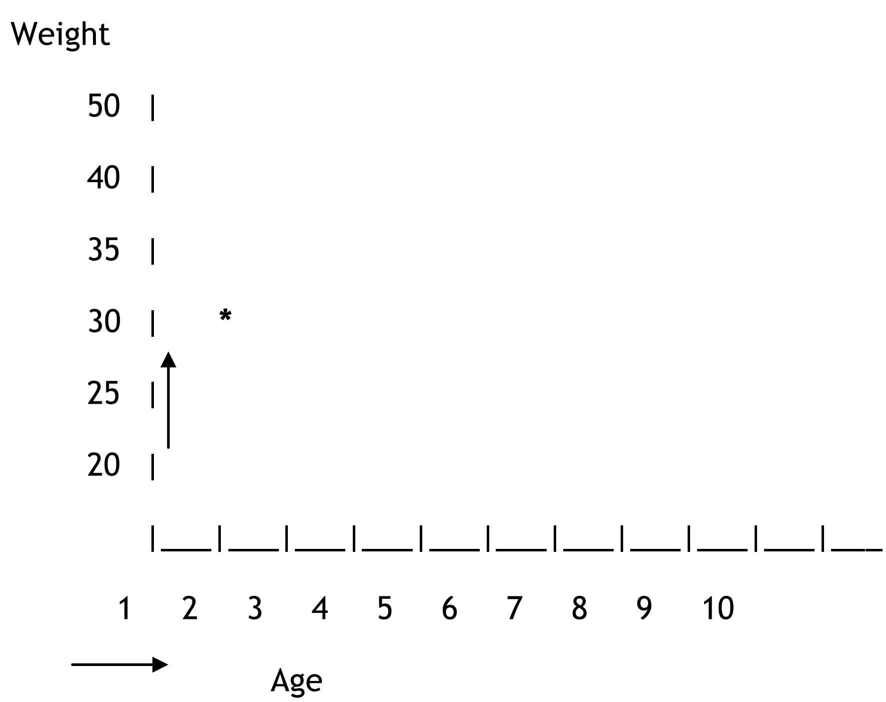
Null hypothesis. See "Hypothesis.

Objective measures and measurement. Some things are not objects, in the sense that they can be directly seen, heard, and touched. Examples include attitudes and feelings. These are **subjective**---known by the **subject**, the person. Other things are **objective**. They can be directly seen, heard, and touched. Therefore, unlike nonobjective/subjective things, they are "available" to be observed by multiple persons. Examples include behavior (such as the number of math problems students solve) and interaction (such as the number of times students correctly answer questions and the number of times the teacher provides specific praise for correct answers. Things that are objective can be counted; i.e., there can be quantitative measurement and information.

If a thing is objective, it is best to measure it quantitatively---to count it. To merely summarize it with an opinion ("I think students know letter-sound relationships very well") provides less precise information than summarizing the same thing with a quantitative statement such as, "15 out of 20 students give the correct sound to the letters 95 percent of the time." This information can be used to make decisions. The qualitative (subjective) statement cannot.

Observational research. This is used to collect information on ongoing actions and interactions; e.g., student-teacher interaction, students' behavior on the playground, student strategies for conducting experiments. Data are usually collected through direct observation, either in a narrative (sportscaster form) or by scoring pre-formed recording sheets (e.g., the observer scores whether the teacher provided timely error correction each time an error occurred in a lesson).

Plot data on a graph. A graph or chart usually has two lines: one for each of two variables. For example, the bottom line (across) might be time in years (1 year old, 2 years old, etc.) And the up line might be weight in pounds.



You have a sample of children of different ages. You know each child's age and weight. To plot data on each child, you find the child's age on the across line (say, 2 years) and then move up to until you get to weight (say, 30 pounds). You put a dot of some kind at the spot that shows 2 years/30 pounds.

Pool. A pool is the set of persons, classrooms, schools, districts, states, nations from which you draw a sample. The pool may not be the entire population.

Population. A population is the total set of persons, classrooms, schools, districts, states, nations that have characteristics that you wish to measure. For example, the population of all students who received a new reading curriculum for one year.

Post-test only design. This is an experimental design in which no pre-test is given. If there is no comparison group, it is largely useless as a way to determine effectiveness, because you have no way to tell where a group began. However, if you have equivalent experimental and control groups, it may be assumed (very tentatively) that their pre-test scores were probably similar. Therefore, if the experimental group's outcome scores are significantly different from the control group's outcomes scores, there is reason to suspect (but not to be convinced) that the intervention made the difference.

Pre-test, post-test design. This is a kind of experiment in which data are taken (for example, on students' math skill) before and at the end of an "intervention," a teaching method is used, or a change is made in class. If nothing else changed between the pre-test and post-test (except the delivery of instruction), then it is likely that any increase in students' knowledge (shown by comparing the pre-test scores and post-test score) is the result of instruction.

A pre-test, post-test design with one group is not as powerful as a pre-test, post-test design that uses an experimental group and control group. If you have only one group, other factors COULD have operated between the pre-test and the post-test that affected post-test scores. For example, some children got tutoring, and that made their scores higher. If the researcher concludes that the class scores were higher at the post-test BECAUSE of the new math curriculum, this claim would be Invalid.

The experimental design that has an experimental and control group means that any OTHER changes in the groups between the pre-test and the post-test (e.g., tutoring) could have happened to both groups. Therefore, the ONE main difference is STILL the difference in curriculum.

Purposive sample. If you use simple random sampling, you may not obtain in your sample persons, groups, classrooms, schools, etc., that have characteristics that are relatively rare. Therefore, you would purposively sample (find) persons, groups, etc., for your sample.

Qualitative data. Qualitative data are opinions, perceptions, interpretations. They are answers to questions such as, "How would you describe your students' effort overall?" Qualitative data help to complete the picture provided by numbers---quantitative data. Because they are so subjective, qualitative data should not be used to judge the effectiveness of a curriculum or teaching method---any more than feeling a person's arm should be used to measure blood pressure.

Qualitative research. Qualitative research gains information on the opinions, beliefs, and interpretations of persons in a social environment in order to better understand how persons make sense of their activities. Qualitative research collects information through direct observation (with an emphasis on conversations and physical aspects of an environment that may reflect persons' perceptions; e.g., what does it mean if low performing students are in the back of the class?) and informal interviews. Qualitative research may complement quantitative research and quantitative data. For example, quantitative research on student achievement (does a new math program raise student achievement more than the current math program) may be complemented by qualitative data on how teachers feel about the new math program and their students' interest, effort, and achievement.

Quantitative data. Numerical data, such as scores on tests, percentile rank, percentage of students who are graduated from high school. Quantitative data provide more precision than qualitative data. (See Levels of measurement.)

Quantitative research. Quantitative research gains information on the values (how much there is) of variables that have been identified. It is generally used to describe: (1) the **current levels** of variables (such as students' scores on achievement tests; the percentage of students who are affluent, middle class, and poor; the percentage of minority and nonminority students; and numerical data on teacher behavior, such as the percentage of times teachers correct errors); and (2) **changes (e.g., between pre-test and post-test) in the levels of variables** after an intervention (e.g., a new reading program is introduced; teachers receive special training).

Randomization, or random allocation. This is a second way, besides matching, to try to make experimental and control groups equivalent. If you have a "pool" of 50 students, you randomly assign them to the two groups. This means that all factors (ethnicity, social class, family support, background knowledge, age, sex) have an equal chance of being in either group.

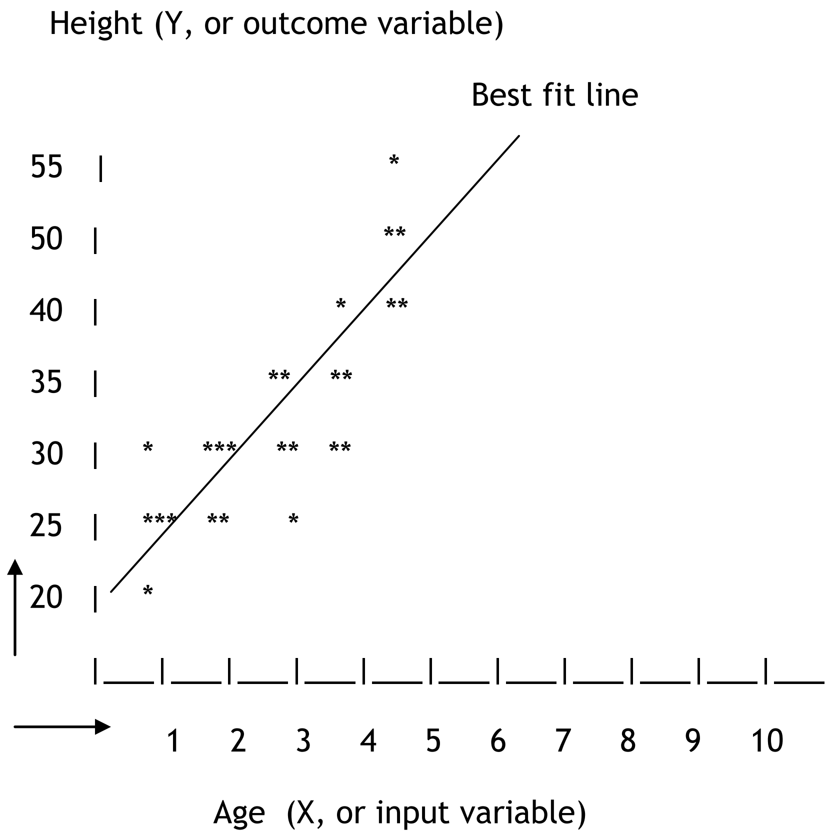
Range. Range means the spread of scores from lowest to highest. For example, a group of persons ran as far as they could. The shortest distance run was 1 mile. The longest distance was 40 miles. So the range is from 1 to 40 miles. It doesn't matter if only one person ran 40 miles or if five persons did. Range is not interested in how many. It is only interested in the spread.

Reliability. Reliability means repeatability. If two different observers or testers obtain the same scores on the same thing, then the scores are reliable. If the findings from the same research conducted with different persons or schools are much the same, then the findings are reliable and the instruction that produced the same findings (e.g., student achievement) is said to have reliable effects.

Replication. Replication means that the research is conducted again and again with the SAME samples, to see if the results (e.g., of a new curriculum) are reliable. If so, then it is NOT likely that the results of the first study were a fluke of some kind. Replication also means that the same research is conducted with DIFFERENT samples. This enables researchers to find out if an "intervention" (e.g., curriculum, teaching method, classroom routine) works better in certain situations. It is a way to determine the GENERLIZABILITY of findings.

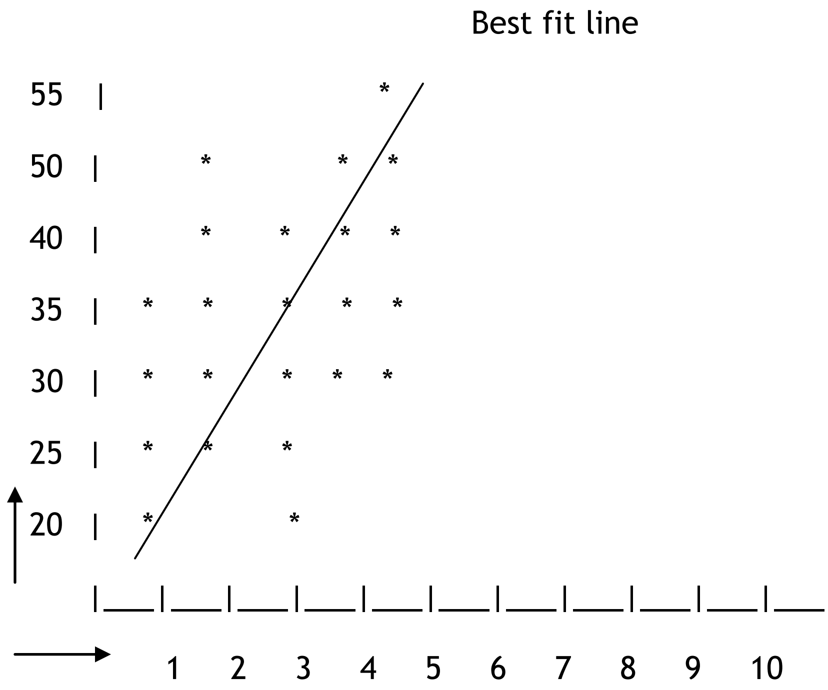
Representative Sample. A sample whose characteristics (e.g., percentage of persons of different ethnicities, social classes, sexes, skills) are similar to the characteristics of the population to which the findings are relevant. If a research sample is not representative of the relevant population, you cannot logically claim that the findings can be generalized to the population.

Relationship (association, correlation). A relationship or correlation means that values of one variable (usually the X or input variable) predict values of another variable (usually the Y or outcome variable). It is rare that values of X predict values of Y exactly. Instead, values of X may predict a range of Y values. For instance, there is obviously a correlation or relationship between age and height. The older persons are (up to a point---such as age 21) the taller they are. But each value of X (age, such as 1, 2, 3, and 4 years old) does not predict height exactly. Instead, each age predicts a range. For instance, a sample of 25 children RANGING in age from 1 to 4 years, may show that the five children who are five years old range from 40 to 55 inches. So, knowing the value of age (5 years old) predicts anywhere from 40 to 50 inches. This is not exact.



Now let’s draw a best fit line THROUGH the plotted data points. each of which represents a value of X and corresponding values of Y. The best fit line is drawn THROUGH the dots in such a way as to minimize the distance between the line and the dots. The line shows the TREND. As you can see, the trend is upward. That is, the relationship between age and height is a DIRECT relationship. The older the age, the greater the height. As age changes, height changes in the SAME DIRECTION. How close are the dots to the best fit line? Fairly close. This means that the relationship between age and height is fairly strong. But look at the data below.

Height (Y, or outcome variable)



Age (X, or input variable)

This time, the range of values of Y (at each value of X) is much greater. In other words, X does NOT predict Y as accurately as it does in the first graph, above. Compare visually the distances between the data points and the best fitting line in the two graphs. Also compare the range of numerical values. The visual distances and the numerical ranges are larger in the second graph. In other words, even though there is a trend (older is USUALLY TALLER) X is a WEAK predictor of HOW tall. We would call this a LOW correlation.

Scientific reasoning. The use of objective data to test beliefs and draw conclusions about the truth or accuracy of the beliefs. OBJECTIVE EVIDENCE, not on opinions or beliefs. Generally, instances (e.g., groups, schools) that have one feature are **compared and contrasted** with otherwise similar instances that do NOT have the feature. Data are collected to see if there are any OTHER differences that can account for the main one. For example, one group of persons with arthritis is given a new drug. Another group that is similar in age, onset of arthritis, and severity of arthritis is NOT given the new drug. If the group that got the drug (experimental group) improves significantly, and the other group (the control group) that did NOT get the drug does NOT improve much, then drug is the likely reason, or cause of the difference in improvement.

Survey research. This is research designed to gain, literally, an overview. Survey research usually involves selecting a sample and then using interviews and questionnaires to obtain information that describes the big picture. It usually provides information on how “things” are or how they have changed, but it does not usually involve any efforts (intervention) to test or to effect change. It is most useful for obtaining information on opinions, beliefs, attitudes.

Trend. On a graph, a trend means that there is regular change. The graph below shows data for 21 persons---21 data points. We know the shoe size of each person, and we know how many books each person read last year.

```
Books read per year

50 |
40 |                        *                         *
35 |                                 *          *
30 |             *     *    *    *                  *
25 |      *     *             *             *          *
20 |      *        *          *    *   *         *    *
   |______|______|______|______|______|______|______|______|______|______|______
    1      2      3      4      5      6      7      8      9     10     11

                              Shoe size
```

Is there a trend here? For example, is it the case that the larger the shoe size the more (or less) books a person reads? NO. Persons with a size 1 shoe read 20 and 25 books. But persons with a size 10 shoe ALSO read 20 and 25 books.

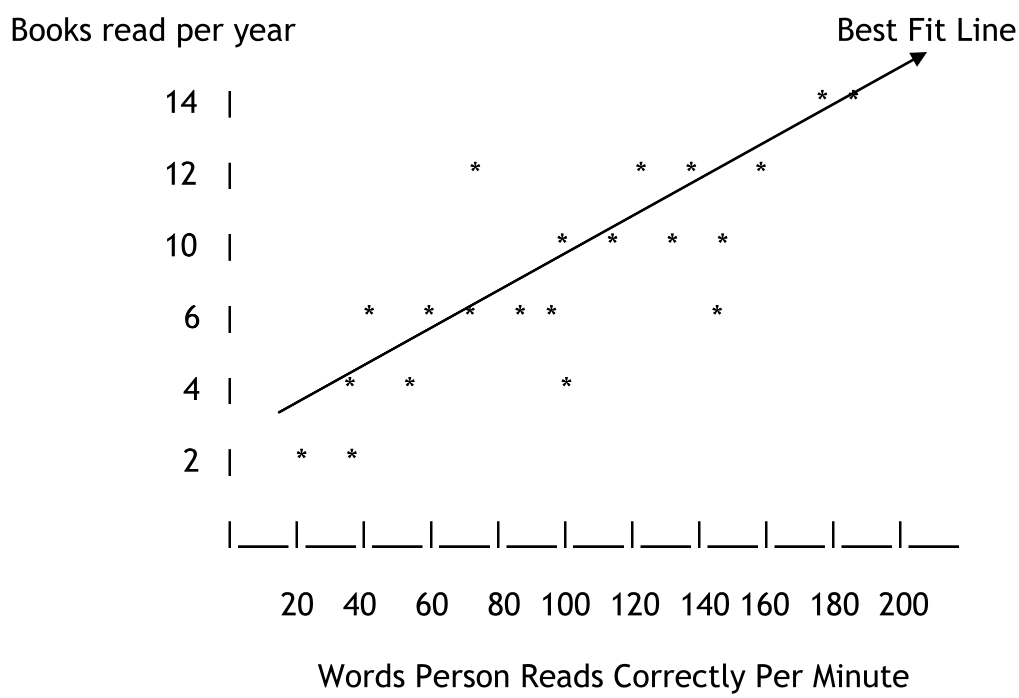
Here's another graph.

```
Books read per year

14  |                                  * *
12  |             *         *  *   *
10  |                   *  *  *  *
 6  |       *   * *  * *         *
 4  |      *   *        *
 2  |   *  *
    |___|___|___|___|___|___|___|___|___|___|___
        20  40  60  80 100 120 140 160 180 200

          Words a Person Reads Correctly Per Minute
```

It shows data for 21 teenagers. We know two things about each person: how many books they read last year and how many words they read correctly per minute (reading fluency). So, if you look at the bottom left corner, it PLOTS the data for one person. He reads 20 correct words per minute (very slow) and he read 2 books in a year.

Now look at the right side of the graph. Two persons read at a rate of 200 correct words per minute; one read 12 books and the other read 14 books.

Do you see a trend? For example, does the number of books per year change as the fluency increases? Yes.



The best fit line does NOT connect the plotted data points. It cuts through them so that there are about as many above it as below it.

Triangulation. Using multiple measures of the same thing (variable). If different kinds of data (e.g., questionnaire, test scores, classroom observations) all say the same thing (e.g., the teacher is competent), then the finding is likely to be more valid (accurate, representative of the facts) than only one source of data.

Validity. Validity generally means that statements accurately represent what IS: the facts. There are several uses of the word validity in research.

- The extent to which an instrument or single measure in fact ***measures what it says it measures***. For example, how a child holds a book is **not** a measure of (is not an example of) reading. But how many words a child accurately reads per minute **is** ONE measure (example of) reading. This kind of validity hinges on **definitions**.
- Validity is also the extent to **which findings accurately represent** what in fact happened. For example, if a researcher reported that the average number of correct answers on a test was 75, but in fact the average was 65, the finding is not valid. This kind of validity hinges on accurate **measurement and reporting**.
- Validity is also the extent to which **claims are supported by hard evidence**. For example, if a writer says that teachers should adapt instruction to students' learning styles, and in fact there is no experimental evidence, or no credible (believable) experimental evidence to support this claim (more than the opposite claim---that it makes little difference if teachers adapt instruction to students' learning styles), then the claims are not valid. This kind of validity hinges on all aspects of research: definitions of variables (what is a learning style? How do you know a person has a certain learning style?); and how you tested the HYPOTHESIS that adapting instruction to students' learning styles makes a difference.

Variable. A variable is any KIND of thing that is part of a description or explanation. Another name might be "factor." Variables have different values. They vary in value. For example, weight is a variable. One person's weight in 150 pounds. Another person's weight is 250 pounds. One school's achievement rate (a variable) is 90% of students read above grade level. Another school's achievement rate is 75% of students read above grade level.

Variables differ in the part they play in an explanation. For instance, here is how we might explain achievement. Following is out CAUSAL MODEL.

Quality of	→	[Given the degree of	→	Student achievement
curriculum		teacher proficiency using		
materials		the materials]		

Student achievement is seen as an OUTCOME variable. An effect. A dependent variable. It is seen as an outcome or effect that is dependent up curriculum and teacher proficiency.

The quality of the curriculum can range from poor to excellent. Quality of curriculum is seen as a main INPUT variable. A cause or predictor of achievement. An Independent variable.

Notice that the quality of the curriculum does not by itself cause or predict achievement. It depends on something else---namely, the proficiency of the teacher. We would call teaching proficiency an INTERVENING variable. It intervenes or comes between the main independent variable (curriculum) and the main dependent variable (achievement). Therefore, our model shows that an excellent curriculum will NOT produce or predict high achievement unless teaching proficiency is also high.