Scientific Thinking

Martin Kozloff

Let's begin with a few….

Cautionary Tales

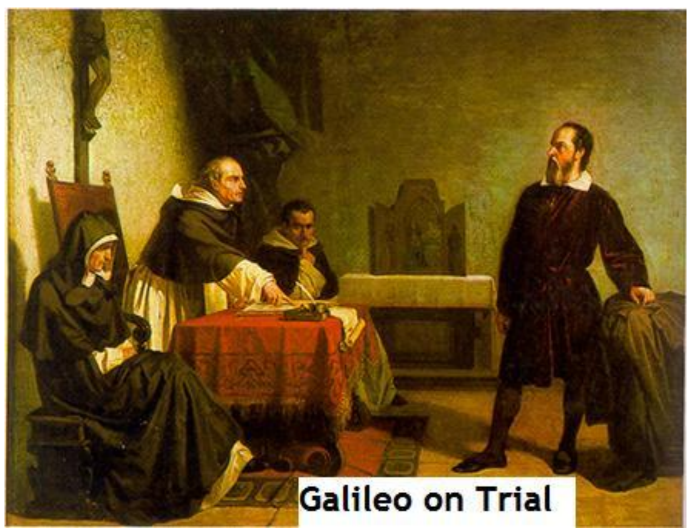
"…AND YOU'LL KNOW THE TRUTH, AND THE TRUTH WILL SET YOU FREE."

[John 8:32]

Yeah, well, maybe.

Cautionary Tale 1. Galileo ("Seeing is believing.") versus The Church ("Believing is seeing.")

Think about that. You get it? Okay, what's it mean?



Galileo to Kepler, 1610

"My dear Kepler," says Galileo, "what would you say of the learned here, who, replete with the pertinacity of the asp, have steadfastly refused to cast a glance through the telescope?"

"Just look through the telescope. You'll SEE that I'm right!"

"NO! We already KNOW that you're wrong."

"For Pete's sake, LOOOOOK!"

"NO!"

"What shall we make of this?," Galileo asks Kepler. "Shall we laugh, or shall we cry?"

http://www.law.umkc.edu/faculty/projects/ftrials/galileo/galileo.html

Copernicus theorized that the sun was the center of the universe and that the earth revolved around the sun. Galileo **tested** the Copernican theory with **data** collected through his telescope. Sure enough, Copernicus was right. Galileo published his data. This got him in trouble with the Church, whose doctrine was that the earth----not the sun----was the center of the universe. So, they put Galileo on trial. They found Galileo guilty.

> "The proposition that the Earth isn't the center of the world and immovable but that it moves, and also with a diurnal motion, is equally absurd and false philosophically and theologically considered at least erroneous in faith."

They sentenced Galileo to prison.

> "We condemn you to the formal prison of this Holy office during our pleasure, and by way of salutary penance we enjoin that for three years to come you repeat once a week at the seven penitential Psalms. Reserving to ourselves liberty to moderate, commute or take off, in whole or in part, the aforesaid penalties and penance.
>
> And so we say, pronounce, sentence, declare, ordain, and reserve in this and in any other better way and form which we can and may rightfully employ."

Galileo was eventually vindicated. **But first he died.** The excesses of the Church (suppression of individual research, thinking, and speech; intransigence in the face of contradicting facts) helped to bring in the Reformation---which was not a total blessing. But Galileo's ideas (and empirical methods) won.

What does Galileo's trial have to do with education?

Some persons in this field are **empiricists**---but **not** necessarily so radically empirical that they believe we can know NOTHING except through experience. See also Roger Bacon here and here; and Francis Bacon, on inductive reasoning and on the "idols" that distort knowledge of the truth. Like Galileo, modern day empiricists follow the facts---the data. Either (1) facts lead them to **develop** beliefs ("Okay, so I *think* it works like this....") or (2) they use facts to **TEST** beliefs ("Let's collect data to see **if our hypothesis [beliefs] are *wrong***.").

However, many persons and groups in education (**anti-empiricists**) search for facts that support what **they already believe**. You can always find support. Interview enough persons and someone will agree with you. It's called "cherry picking." Anti-empiricists also **ignore facts that contradict** what they believe. Still other anti-empiricists **collect no data at all**. They know in their hearts that they are right.

> "Phonics is incompatible with a whole language **perspective** on reading and **therefore is rejected**." [Watson, D. (1989). Defining & describing whole language. *Elementary School Journal, 90*, 129-142.] *Sounds like the same illogic spoken at Galileo's trial.*

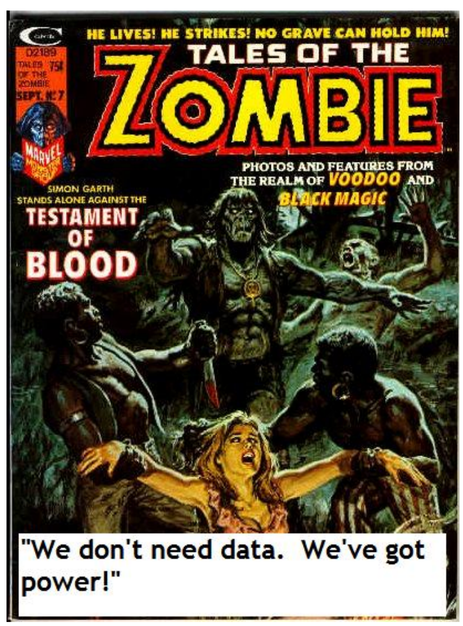
> "It seems futile to try to demonstrate superiority of one teaching method over another by empirical research." [Weaver, C. (1988). *Reading: Progress and practice*. p. 220. Portsmouth, NJ: Heinemann.] *Is it really futile?*

> "***Early*** in our miscue research, ***we concluded***…That a story is easier to read than a page, a page easier to read than a paragraph, a paragraph easier than a sentence, a sentence easier than a word, and a word easier

than a letter. Our research continues to support this conclusion and **we believe it to be true…**" Goodman, K. & Goodman, Y. (1981). *It's easy to do research in a way that supports what you believe? That's why OTHER persons should test what you believe. It's called "independent research."*

BEWARE of anti-empiricists. ***They will sell you a mirage.***

The battle between empiricists and anti-empiricists in education has been fought a long time. The **anti-empiricists** in schools of education, in school districts, and in national curriculum organizations (such as National Council for Teachers of English) have had control for decades. That's why there are so many untested and harmful fads---for example, in reading and math. But now, with No Child Left Behind, Reading First, and state accountability systems, the pendulum may be swinging back to scientific reasoning---rather than unfounded belief, doctrine, and speculation---to make education decisions. But the anti-empiricists are still around. Like zombies, they don't die.



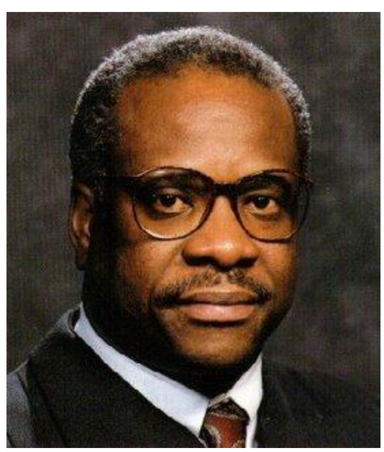
Here's the lesson for persons who want to be leaders.

The truth WILL set **YOU** free----**from error**. But before it helps you to set **other** persons free, you'll be condemned by those whose power (position, control, prestige, and privilege) requires unquestioned acceptance of their doctrine. For, if it's shown that they're wrong on one count, they may be wrong on other counts; therefore, they are fallible; therefore, they must be wrong about many things; therefore, they aren't legitimate authorities and shouldn't be trusted or obeyed.

But if you yield to dictates of the powerful, and to pressures from the herd of believers in order to avoid confrontation and to feel safe, you'll become a coward. And persons (children, teachers, civilization) who depend on you for the truth and for your strength to defend THEM against the herd and the powerful, will be sacrificed on the altar of your fear. [Please click that link.]

If you assert the right to think for yourself---to be skeptical, to require credible data, to challenge the group mind and the dictates of persons in power---you may be ridiculed, threatened, and even lose your job. But if you persist, you may just win. And in the meantime, you'll be serving something more important than your desire for temporary security (bought at the cost of your soul); namely, the truth and your moral obligation not to harm children. Besides, do you want to live on your knees?

Here's what Justice Clarence Thomas has to say on the matter.



> If you trim your sails, you appease those who lack the honesty and decency to disagree on the merits, but prefer to engage in personal

attacks. A good argument diluted to avoid criticism isn't nearly as good as the undiluted argument because we best arrive at truth through a process of honest and vigorous debate. Arguments shouldn't sneak around in disguise, as if dissent were somehow sinister or clandestine. **One shouldn't be cowed by criticism.**

In my humble opinion, those who come to engage in debates of consequence and who challenge accepted wisdom should expect to be treated badly. Nonetheless, they must stand undaunted. That is required. And that should be expected. For it is bravery that is required to secure freedom.

On matters of consequence, reasons and arguments must be of consequence. Therefore, those who choose to engage in such debates must themselves be of consequence. Much emphasis these days is placed on who has the quickest tongue and who looks best on television. There seems to be an obsession with how one looks to others; hence, a proliferation of public relations professionals and spin doctors. As I was counseled some years ago, perceptions are more important than reality. But this is madness. No car has ever crashed into a mirage. No imaginary army has ever invaded a country.

What makes it all worthwhile? What makes it worthwhile is something greater than all of us. There are those things that at one time we all accepted as more important than our comfort or our discomfort - if not our very lives: duty, honor, country. There was a time when all was to be set aside for these. The plow was left idle, the hearth without fire, the homestead abandoned.

http://americanradioworks.publicradio.org/features/sayitplain/cthomas.html

Cautionary Tale 2. Decisions based on Invalid Data

Iris Ledbetter, Principal at Eldorado Elementary, worked with teachers to select math materials that would raise achievement. They chose *Holistic Math*. Staff worked hard, but achievement didn't rise. It fell. The materials were poorly designed. Sadly, neither Ms. Ledbetter nor the teachers knew how to examine the **evaluation research base** for *Holistic Math*---which would have shown that: (1) the researchers used such **broad definitions** of achievement that changes in **NONmath** behaviors (such as student "interest in math") made the materials look good, even though the kids did poorly on many REAL math skills; and (2) the researchers didn't control for, or even consider the effects of **extraneous variables** (such as maturation) that accounted for some of the alleged progress of children in their research. Result? Students at Eldorado got poor math instruction three years in a row. So, it was nearly impossible to succeed with middle school math. **Nice work, Iris!**

Cautionary Tale 3. Decisions Based on a Simplistic Picture of the Causal Process.

Jose' Ramirez, Assistant Principal at Hoarse Coyote High School, planned and implemented remedial reading. He and his teachers examined the evaluation research bases for many remedial reading programs, and wisely selected the one with the most **credible data** showing effectiveness. Mr. Ramirez made sure that teachers in the remedial reading classes were trained to high proficiency and implemented the programs carefully. Yet, **student**

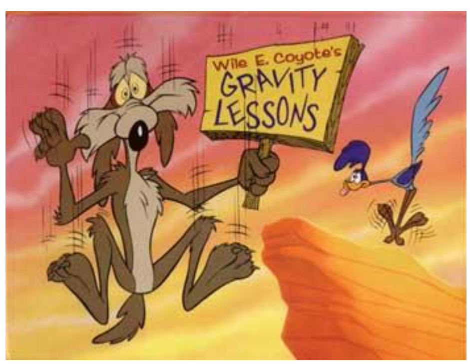
progress was minimal---far below what the research had led Mr. Ramirez to **predict**. Why? Because Mr. Ramirez presumed that the **causal relationship** was like this:

Properly Implement	→	Substantial Reading
Program		Achievement
(Independent		(Dependent
variable)		variable)

In fact, however, the effectiveness of the program ALSO required that teachers in **content** areas (math, history, literature) help students to apply or **generalize** reading skills from their remedial class to **new materials**, and this would have required planning, direction, and supervision (leadership and management). In other words, the causal sequence is **really** like this.

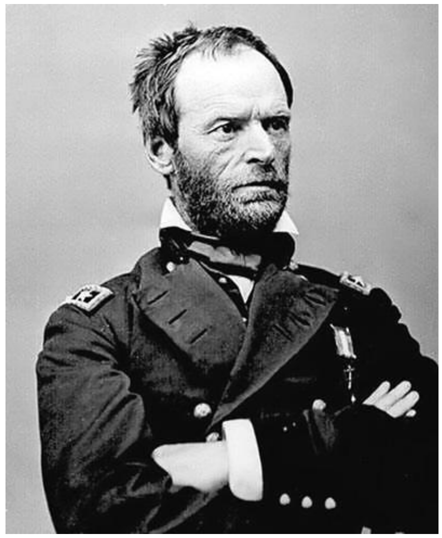
Properly Implement	→	[If Leadership and	→	Generalization	→	Substantial Reading
Program		management.		to Other Classes]		Achievement
(Independent		(Intervening Variables)				(Dependent
variable)						variable)

Too bad, Mr. Ramirez. You're a great guy, but you need to think with **more precision**.



Cautionary Tale 4. Sucker for a Sales-pitch

William Tecumseh Shermanski, Principal at Cannonball Middle School,



was so impressed by presentations at a conference on "Learning and the Brain" that he purchased a new program for his school---*Brain Blow-out*. The materials were cheap, but required lots of planning and instructional time. **Totally wasted!** Teachers began to see Mr. Shermanski as a **poor leader**.

"Gee, like we don't have enough to do without idiotic fads!"

Despite his cool Yankee uniform, his good intentions, and his heroic namesake, Mr. Shermanski was fooled by a sales-pitch that used **evocative** phrases like "*Brain Blow-out* strengthens functions in both hemispheres," "is research based," "involves authentic learning," and "is holistic and natural." Too bad he wasn't buying candles and incense instead of instructional materials.

Do YOU want to be INeffective?

Do you want to WASTE money, time, energy, and teachers' trust?

Do you want people to think you're a NINCOMPOOP?

Or a clown?

No?

WELL, IT'S GONNA HAPPEN IF YOU DON'T KNOW HOW TO THINK SCIENTIFICALLY!

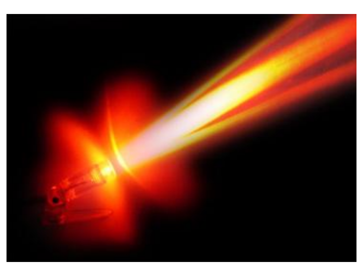
Making sound educational decisions isn't like betting on red 18 in roulette.

If you use **guesswork** to make decisions, or if you make wrong decisions because you're **not thinking clearly**, you'll have a mighty mess on your hands.

"Duuhhheeeee."

You need to think like a laser beam cutting through metal.

Precise. Straight. Powerful.



Like a sniper aiming at a target 1000 yards away.

Like a…. Well, you get the point.

Which leads us to…

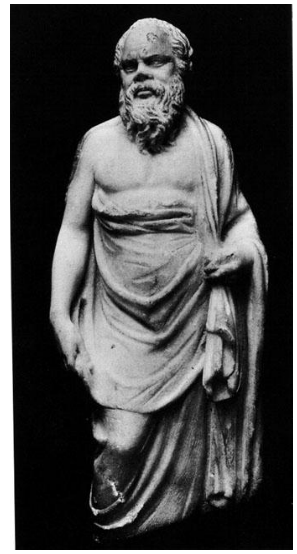
Features of Scientific Thinking
In Contrast to Unsupported Belief, Doctrine, Speculation, Fantasy, and Flapdoodle

Scientific thinking is simply valid reasoning applied to scientific questions. However, the same principles and procedures of valid reasoning can (we hope they are) used in medical diagnosis, selecting curriculum materials, buying a car, and deciding whether to have a face lift.

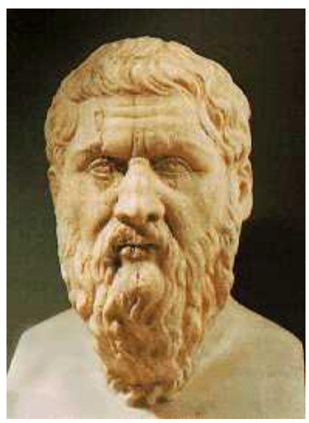
> "Let's see. 20 thousand bucks. Pain. Lasts two years. End up like this lady. Calls herself "Madonna." No, thanks!"

Plato and Definitions

2500 years ago, Plato tried to teach his fellow Greeks how properly to **define concepts** and **test beliefs** by **examining whether the data supported them**. Here's a sample from Book 1 of *Republic*. Socrates (Plato's friend and teacher) is discussing the concept of justice.



One of the group is Thrasymachus, a blowhard and a sophist.



Sophists were like modern-day political advisors. They taught rulers and officials **not** how to learn and present the **truth**, but how to **craft an argument (using fuzzy definitions and emotive phrases) that the public would believe**. Plato and Socrates despised sophists---whose teaching meant that political decisions would be based on falsehoods sold to gullible voters by a clever good [*My comments are in italics.]*

Thrasymachus. Listen, then, I proclaim that **justice isn't anything else than the interest of the stronger.** …

Socrates. Let me first understand you. Justice, as you say, is the interest of the stronger. What, Thrasymachus, is the **meaning** of this? You cannot mean to say that because Polydamas, the pancratiast [*full contact fighting*], is stronger than we are, and finds the eating of beef conducive to his bodily strength, that to eat beef is therefore equally for our good who are weaker than he is, and right and just for us? [*In other words, the definition may be too broad*.]

Thrasymachus. That's abominable of you, Socrates; you take the words in the sense which is most damaging to the argument.

Socrates. Not at all, my good sir; I am trying to understand them; and I wish that you would be a little clearer. [*Sarcasm*]

Thrasymachus. Well, have you never heard that forms of government differ; there are tyrannies, and there are democracies, and there are aristocracies.

Socrates. Yes, I know.

Thrasymachus. And the government is the ruling power in each state?

Socrates. Certainly.

Thrasymachus. And the different forms of government make laws democratical, aristocratical, tyrannical, **with a view** to their several interests; **and these laws, which are made by them for their own interests, are** *[what they consider]* **the justice which they deliver to their subjects**, and him who transgresses them they punish as a breaker of the law, and unjust. And that is what I mean when I say that in all states there is the same principle of justice, which is the interest of the government; and as the government must be supposed to have power, the only reasonable conclusion is, that **everywhere there is one principle of justice, which is the interest of the stronger**.

Socrates. Now I understand you; and whether you are right or not I will try to discover. But let me remark, **that in defining justice you have yourself used the word "interest"** which you forbade me to use. It is true, however, that in your definition the words "of the stronger" are added.

Thrasymachus. A small addition, you must allow.

Socrates. Great or small, never mind about that: **we must first enquire whether what you are saying is *the truth***. Now we are both agreed that justice is interest of some sort, but you go on to say "of the stronger"; **about this addition I am not so sure, and *must therefore consider further*.** *[Socrates will expose the **implications** of this definition, which are contrary to what Thrasymachus asserted. Therefore, the definition must be false.]*

Thrasymachus. Proceed.

Socrates. I will; and first tell me, Do you admit that it is just for subjects to obey their rulers?

Thrasymachus. I do.

Socrates. But are the rulers of states **absolutely infallible**, **or are they sometimes liable to *err***?

Thrasymachus. To be sure, they are liable to err.

Socrates. Then in making their laws they may sometimes make them rightly, and **sometimes not**?

Thrasymachus. True.

Socrates. When they make them rightly, they make them agreeably to their interest; **when they are mistaken, contrary to their interest; you admit that**?

Thrasymachus. Yes.

Socrates. And the laws which they make must be obeyed by their subjects, — and that is what **you call justice**?

Thrasymachus. Doubtless.

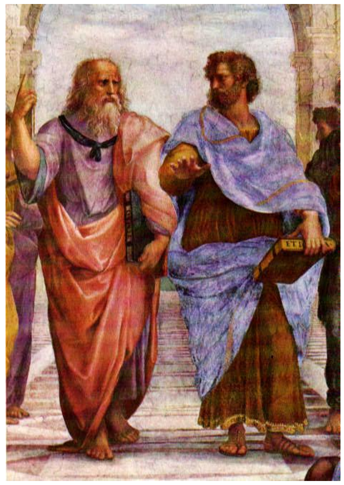
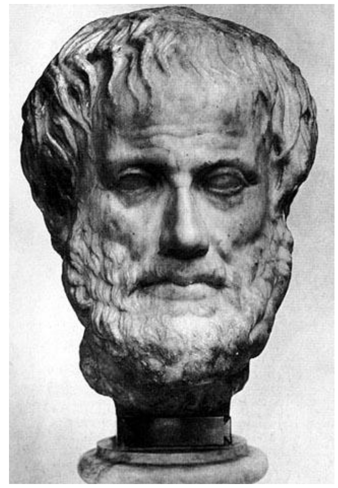
Socrates. Then justice, according to your **argument**, isn't only obedience to the interest of the stronger **but the reverse**?

[Plato. Republic. http://www.constitution.org/pla/repub_01.htm]

Socrates showed that when a definition is too broad (justice is defined as **whatever** the stronger believe to be in their interests), it leads to false conclusions (e.g., If rulers think that something is in their interests, but are wrong, then their "justice" would NOT be in their interests). Therefore, **decisions based on these false conclusions are likely to make a mess of things**.

Aristotle on Reasoning

Following Plato (who was his teacher), Aristotle wrote many books on logical thinking. http://vos.ucsb.edu/browse.asp?id=258

Plato (left) and Aristotle

Here's a selection from *Prior analytics*.

> A premiss [*or premise*] then is a sentence affirming or denying one thing of another. This is either universal [*all or none*] or particular [*some*] or indefinite. By universal I mean the statement that something belongs to all or none of something else;
> *["All cats are felines." "No Persians can be trusted."]*
> by particular that it belongs to some or not to some or not to all;
> *["Some birds cannot fly."]*
> by indefinite that it does or does not belong, without any mark to show whether it is universal or particular…
> *["Dogs have four legs." ALL dogs?]*
>
> A **syllogism is discourse** in which, certain things being stated, something **other than** what is stated **follows of necessity** from their being so. I mean by the last phrase that they [*the things being stated*] produce the **consequence** [*conclusion*], and by this, that no further term is required from without [*no other data are needed*] in order to make the consequence [*logically*] necessary.
> *[All beings are mortal. Premiss.*
> *Socrates is a being. Fact*

Therefore, Socrates is mortal. Consequence that follows necessarily.]
[Aristotle. Prior analytics. http://classics.mit.edu/Aristotle/prior.1.i.html]

Scientific thinking often is organized as follows.

Theory/beliefs → Variables or Models of Relationships → Definitions →
Measures → Findings → Conclusions → Implications (e.g., decisions)

Here are some principles of scientific thinking.

1. A scientific thinker (in contrast to salesmen or dogmatizer) lays out a theory of the thing studied; e.g., reading achievement. What is reading? What is reading achievement? What variables affect reading achievement?

2. The theory is consistent with sound scientific research (that is, with other validated beliefs). It's not a fantasy. [Can you identify ***whole*** methods of instruction that are based on a phantasmagoric "theory"?]

3. The theory identifies hypothetical independent (input), dependent (outcome), intervening variables, and contributing conditions.

4. A scientific thinker provides conceptual (general) definitions and operational (precise) definitions for important independent (input), dependent (outcome), and intervening variables, and contributing conditions.

5. These definitions use clear and concrete terms. Therefore, the theory refers to the real world. It can be tested.

7. A scientific thinker evaluates the accuracy of a theory (or hypotheses within the theory) by testing the **null hypothesis.** That is, by trying to collect evidence (data) that show that the theory or hypotheses are **false**.

8. Measures are consistent with the definitions of the variables. **Measures measure what they purport to measure**, and not something else. Measuring reading proficiency in terms of how many books children read is invalid. How many books children read may be an **effect** of reading skill, but it isn't a **definition** of reading skill.

9. Measurement is direct. When possible, observe, don't ask.

10. A scientific thinker measures variables at the proper *level* or *scale* of measurement: nominal, ordinal, interval, ratio. If you can count it (ratio level), then count it. If it isn't the sort of thing that can be counted (e.g., how much persons agree with a statement---ordinal level) then don't count it. Use an ordinal scale.

    Self-confidence: Very little Fair amount A lot Great deal

11. When possible, a scientific thinker uses several measures of the same variables---triangulation.

    Measure 1 Measure 2 Measure 3

    ↘ ↓ ↙

    Thing (variable) measured

12. A scientific thinker assesses and reports the reliability of measurement.

13. A scientific thinker summarizes numerical data with range, mean, median, and mode. For example, the average (mean) score may be 23. But what if scores range from 2 to 90? How well does the mean represent the scores? Not well. Therefore, you also want to know the **mode**---the most frequent score. And the **median**—the score for which half the scores are above and half the scores are below. The median might be 40.

"On average, patients recover in five weeks"
Great, but what if the weeks range from 2 to 30?

14. A scientific thinker reports raw numbers as well as percentages. Otherwise, a big **percentage** difference may conceal small actual differences. "33% more projects were done by cooperative learning groups." In fact, one group did 3 projects, and the other group did 4. One more than 3 is 33%. 33% more means ONE. Is this a big deal?

15. A scientific thinker uses proper statistical tests to report the **significance of differences** between groups or between pre-tests and post-tests.

    There may be a 10 point difference in achievement between a pre-test and a post test. You might think this is a big deal. But if there are only 20 students in the class, you could easily get this difference by **chance**!

16. A scientific thinker determines the **degree of association or correlation** between variables in causal research; e.g., comprehension scores as a function of children's fluency scores. What if 10 different fluency scores are associated with one comprehension score? And one fluency score is associated with 10 different comprehension score? How strongly does fluency predict comprehension? NOT.

17. A scientific thinker considers and designs research to test both the research hypotheses and **rival hypotheses** that might weaken internal and external validity; e.g., measurement error, maturation.

18. A scientific thinker draws conclusions that are consistent with the findings. Nothing here. Weak. Possibly something here. Modestly strong. Very strong. The scientific thinker tests causal hypotheses with four questions.

(1) Is there valid and reliable data on both the alleged cause and the alleged effect?

(2) Did the alleged cause precede the alleged effect? Observed? Common sense?

(3) Is there logical evidence (via inductive reasoning) that the alleged cause and effect are connected? [p. 88]

a. Method of concomitant variation.

b. Method of difference.

c. Method of agreement.

d. Method of residues.

(4) Has the possible influence of extraneous variables (threats to internal and external validity) been ruled out or weakened? [p. 98]

19. A scientific thinker replicates the research to see if the findings are obtained again with similar samples (**reliability** of findings) and with samples of different composition (**generality** of findings).

The same logic as getting a second and third opinion.

If you read research reports or listen to presentations that don't follow the above guidelines, then you have NO idea how the researcher ended up with the findings. And it's possible that the researcher doesn't know, either. **You cannot trust the claims.** The proper response is, "Oh, iz zat so?"

Let's begin at the foundation. What does it mean to know something? What is the something that is known. How does "it" come to be known? How is what is known used?

What is Knowledge?

The traditional purpose of education in democratic republics is to pass on and to increase knowledge from one generation to the next in order to sustain the civilization by inculcating civic virtues, reason, and knowledge needed to participate effectively and wisely.

> "I think by far the most important bill in our whole code, is that for the diffusion of knowledge among the people. **No other sure foundation can be devised for the preservation of freedom and happiness... [**Thomas Jefferson to George Wythe, 1786.]

> "In a republican nation whose citizens are to be led by reason and persuasion and not by force, **the art of reasoning becomes of first importance."** [Thomas Jefferson to David Harding, 1824.]

In contrast,

> Just as the purpose of education in monarchies is to enoble men's hearts, so its purpose in despotic states is to debase them. In despotic states education must be servile…. **Absolute obedience presupposes ignorance in the person who obeys**; ignorance is presupposed as well in the person who commands. For he need not deliberate, doubt, or reason; **he has only to will**...Thus education is in one sense nonexistent. Everything previously known must be wiped out, so that something may be taught.
> [Montesquieu. *Spirit of the laws*. 1748]

But what IS knowledge?

Someone says, "Nancy knows English literature." Or, "Ted knows how to solve equations with one unknown, such as y = x – 4. ***What*** do you know when you know English literature or know how to solve equations? **What is IT that's known? The answer is: *connections*.** That sounds weird. But wait. You'll see **that knowledge is *of how things are* connected.** Sometimes we represent, store, and communicate knowledge (of how things are connected) with **language**---spoken or written. For example, a textbook says:

"Oxygen has eight electrons."

This statement **connects** oxygen and electrons. Please read the statement again and notice the connection that it makes. This statement **represents** what we know. This knowledge can be **stored** in books and electronic memories, and it can be **communicated** (taught).

A teacher says, "The line---'She has eyes like a hawk'---is a simile." This statement connects "She has eyes like a hawk" with a literary concept, simile.

So, knowledge (of how things are connected) *can* be contained and expressed in one sentence. But knowledge (of how things are connected) sometimes takes **many sentences**----for example, an **explanation** of (knowledge of) the Revolutionary period in America. Here are some websites where knowledge is stored.

http://www.nps.gov/archive/cowp/Timeline.htm

http://www.historyplace.com/unitedstates/revolution/

http://www.42explore2.com/revolt.htm

http://www.multied.com/Revolt/index.html

http://myrevolutionarywar.com/index.htm

http://www.geocities.com/Athens/delphi/4393/rev_war.html

Please skim a few of them to see **what** is stored. This WHAT (facts, lists, concepts, rules, and routines) would be part of an explanation of the American Revolution?

The explanation (knowledge of how the Revolution started, progressed, and ended) is **a large set of *connected* statements.** It takes many thousands of pages. These statements **make connections** between **persons** (e.g., Samuel Adams, Thomas Jefferson, King George of England, George Washington, Benjamin Franklin), **groups** (Whigs, Tories, Sons of Liberty, Federalists, Anti-federalists), **places** (Philadelphia, Lexington, Boston, New York, Virginia), **events and dates** (Boston Massacre, Sugar Act, Intolerable Acts, Battles at Lexington and Concord), **documents** (*Declaration of Independence*, *Articles of Confederation*, *Constitution*), **cultural ideas and values** (the legitimate role of government, natural rights), and **technology** (printing press, weapons). **So,**

***knowing* the American Revolution would mean *knowing these connections and then* connecting them *into a coherent (connected) story*.** [Please read that last sentence again.] Again,

Knowledge is *of* how things are connected.

Sometimes, these connections are stated or stored in the form of language.

However, *sometimes knowledge isn't in the form of language*. For example, if you make a sharp turn on ice, your car will skid. You know this. You know the connection between turning, surfaces, and skidding. You know the general RULE that connects these things. But when you make a turn, you usually **don't say** this rule knowledge to yourself. You merely **use it**. But you ***could*** say the rule (what you **know**)---to remind yourself. "Okay, it's icy. Slow down. Don't want to skid!" Or to teach a new driver----"Here comes a turn. It's icy. If you turn too fast, you'll skid. So, ease off the gas."

Knowledge is of how things are connected.

Sometimes, these connections are stated or stored in the form of language.

Sometimes they aren't. But they CAN be.

Kinds or Forms of Knowledge [See also *Forms of Knowledge* *for details*.]

By "kinds of knowledge" we don't mean subjects, such as math, biology, and history. We mean something that's the **foundation** of ***any*** subject matter. Remember, knowledge is of how things are connected. **There are different kinds of connections---and therefore, different kinds of knowledge.** Here's a quick introduction, based on the work of Engelmann and Carnine (1991) and Kame'enui and Simmons (1990).

1. **Simple facts. *Facts are knowledge of connections among names, places, and events*.** For example:
    - Boston is the capital of Massachusetts. [The name of a city is connected to the name of a state.]
    - The main export of Saudi Arabia is oil. [The name of a country is connected to the name of a thing.]

- Uranium is radioactive. [The name of a substance is connected to a feature that some substances have.]

Notice that the above facts (a kind of knowledge, a kind of connection) are IN different subjects, or knowledge systems. **All knowledge systems contain facts.** Therefore, facts are one thing you'll teach. There is an effective procedure for teaching facts.

2. **Lists.** ***Lists are knowledge of a string of things that go together***. For example:

- The list of the six New England states.
- The list of the American Presidents.
- The list of the elements in sugar. Carbon, hydrogen, oxygen.

Again, lists---a kind of knowledge---are in all knowledge systems. There is an effective procedure for teaching lists.

3. **Concepts. There are two kinds of concepts—sensory and higher-order.**
   a. **Sensory, or basic concepts.** ***A sensory, or basic concept is one whose features can be seen, smelled, touched, tasted, or heard all at once***. The word "blue" is the **NAME** of a color **concept**. The concept (blueness) is like a circle that surrounds (connects) all examples of (things that have) blueness. Other sensory concepts are in, on, loud, sweet, hard, rough, and circle. Think of a few more. [Go ahead, we'll wait…..]

   You can teach sensory, or basic concepts simply **by showing (and naming) a range of examples** that are **different** in **un**essential features (size, shape) but that all have the common, essential features---they are all BLUE, UNDER, or ROUGH. "This (small, round shape) is blue. This (small, square shape) is blue. This (large, square shape) is blue." Since they are different in some ways, but are the same in ONE way, and are all CALLED the same thing, what do you think the person figures "blue" MUST mean?

   b. **Higher-order concepts. These are abstract concepts.** You **can't** see, hear, touch, or smell their features all a once, as you can with blueness or roughness. Examples of higher order concepts are the **things referred to** by the words (names) democracy, capitalism, courage, or family. **The features of (that define) these concepts are many and they are spread out.** The concept captures or connects all of these features into a "kind of thing." [Please read that sentence again.] However, you can't hold up the examples of higher-order concepts (such as capitalism) as you can hold up a red ball (and say "This is red.") or spray the air with lemon oil and say "This fragrance is citrus." **So, to teach higher-order concepts you give a verbal definition that lists the common features** ("Capitalism is a form of economic system in which there are free markets for goods and services, regulated by the demand, supply, and price.") **and *then* give examples that fit the definition.**

Think of higher-order concepts in different subjects, or knowledge systems.

Literature

Biology

History

4. **Rule-relationships.** ***Rules (sometimes called propositions) are knowledge of how whole sets of things (concepts) are connected.***

- When demand for products (a concept) increases, the price of the product (another concept) increases. [Not demand for one specific item, but demand for products in general. The whole set of things (products) for which there is demand (the concept of demand) is connected to the whole set of things that are prices (the concept of prices).]
- Frogs (one concept) are amphibians (another concept). [Not ONE frog. The whole set of things that are frogs (one concept) is *in* the whole set of things that are amphibians (a second concept).
- As pressure of a gas increases in a closed vessel, temperature of the gas increases. Again, this rule or proposition states what we know of the

connection between pressure of gas in a closed vessel (concept) and temperature of gas in a closed vessel (concept).

Teach rules by (1) stating the rule and then showing examples, or by (2) showing examples and helping students to find the connection and then to state it.

5. **Routines.** ***A routine is a sequence of steps/actions leading to a planned outcome***. Examples include solving equations, adding and subtracting numbers, doing an internet search, writing an essay, brushing your teeth, getting ready for work, or analyzing an historical document. What is connected in a routine? Well, *each step in a routine is an action*: picking up your toothbrush and then loading it with paste (two steps in the teeth-brushing routine), multiplying the numbers in the ones column and then writing the product (two steps in the multiplication routine). **The routine or sequence connects all of the steps/actions into a whole that starts with one action and ends with the outcome.**

  34
x12

Please write the sequence of steps in the routine for two-digit multiplication without renaming (carrying).

First,

Second,

Notice how **the sequence of "First… Second…Third step…" connects the separate steps/actions into a whole**. Also notice that this routine is **general**---you use it to multiply **all** two-digit numbers that are positive and that don't require renaming.

****

Assignment 1. [For your convenience, all assignments are also at the end of this document.]

Examine an historical document, http://www.yale.edu/lawweb/avalon/avalon.htm a textbook, or instruction on the internet. Or go here…

http://people.uncw.edu/kozloffm/knowledge.doc

Identify and write at least two examples of

1. Simple facts.
2. Lists.
3. Higher order concepts.
4. Rules or propositions.
5. Routines. Note that a theory or logical argument would quality as a routine.

****

Knowledge systems

From the time we came into existence, human beings have been trying to figure out what's going on.

"What's **that** stuff?"

"Why does it get light and then dark and then get light again?"

"How can I move this big log from here to over there?"

"Where did this little bitty slimy person come from?"

1. **We've organized what we know** (facts, lists, concepts, rules, routines) about the universe, ourselves, what stuff is made of, and how societies work, **into bodies of knowledge** (astronomy, biology, chemistry, sociology, history, and more).
2. We've **updated** (added to, reorganized) bodies of knowledge as we've learned more---through observation and experiment.
3. And we've **represented**, **stored** and **passed on** this knowledge in stories, poems, and plays (for example, the *Odyssey*, *Beowulf*, *King Lear*), research papers, conversations, lectures, books, and schooling. **So, let's**

call these bodies of knowledge "knowledge systems" (Engelmann and Carnine, 1972).

****

Assignment 2.

Get a textbook in some subject, or a book written on a subject---a knowledge system. Select a paragraph or two. **Identify important things to teach and list them. State what kind of knowledge each "thing" is?** There may be facts and lists; higher-order concepts that the book defines or that you'll have to define for your students; rules stating how one set of things goes with (e.g., causes) another set of things; and maybe routines, such as a theory (a sequence of statements that lead to an explanation). **Do this with a second subject, or knowledge system.** You'll find the same thing. *Everything you want students to learn will boil down to the five kinds of knowledge*.

****

***Acquiring* New Knowledge and Using (*Applying*) Earlier Knowledge are Matters of Logical Reasoning.**

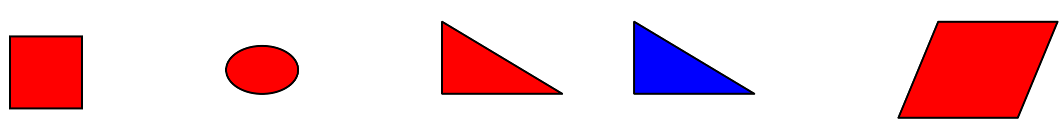
We either (1) learn something new---a new connection---for example, a new concept, rule, or routine; or (2) we apply knowledge (of connections) to something new. **Learning something new involves *inductive* reasoning**---even if we don't know that we are using inductive reasoning. [We USE rules of grammar when we speak and write, but usually don't know THAT we are using the rules. In fact, we usually can't say what the rules **are**!] The following introduces how *human beings use **inductive** reasoning to learn something **new**---to learn a new connection*. The big idea is that

Inductive reasoning begins with specifics or examples of a concept, rule, or routine.

Here are examples.

Mr. Codswallop holds up objects and names each one.

"This is red." "This is red." "This is red." "This isn't red." "This IS red."



The objects called "red" differ in several ways but are the same in one way. The object that is called "NOT read" is **missing the one feature that all of the "red" objects have.** So, what feature of the "red" objects is red? What is missing in the "NOT red" object that makes it not red? ***Why does juxtaposing the red and blue triangles make the difference (that makes the difference) easy to grasp, figure out, see, or induce?***

Ms. Gonzales points to an example of the letter m and says to the class:
"This [m] says mmmmm."
"This [m] says mmmmm."
"This [m] says mmmmm."
"This [m] says mmmmmm."
What do these examples of m have in common that "makes" them say mmmm?

Inductive begins with examples and ends with learners "getting" the general idea (knowledge—the concept, rule, or routine) that *connects* the examples.

Students make generalizations from the examples. They "get" knowledge. "Ah HA! I get it! The color (not the shape or size of the objects) is why the objects are called 'red'" And "The shape (not size, color, or boldface) is what makes the letters say mmmmmm."

Here's another instance of **inductive reasoning**---*going from examples to a general idea that connects the examples*. Let's say that each example below is a school. The left column lists some of each school's features beginning at the **start** of the year. The right column lists certain things that happened by the **end** of the year. We might call these **outcomes**.

Features of Schools Throughout Year	Features of Schools at the End of the Year (Outcomes)
School 1. Small size	School 1. Few suspensions and detentions

Old building High quality instruction High percentage of minority and poor students High quality curriculum Few computers	Added many books to library High student achievement Hired new Assistant Principal
School 2. Small size New building Fair quality instruction High percentage of minority and poor students Low quality curriculum Many computers	School 2. Many suspensions and detentions Added many books to library Low student achievement Five teachers got pregnant
School 3. Large size New building High quality curriculum Low percentage of minority and poor students Few computers High quality instruction	School 3. Few suspensions but many detentions Added few books to library High student achievement No teachers got pregnant
School 4. Large size New building Low quality curriculum Low percentage of minority and poor students Many computers Fair quality instruction	School 4. Few suspensions and detentions Added many books to library Low student achievement Principal retires
School 5. Large size Old building Low quality curriculum High percentage of minority and poor students Few computers Fair quality instruction	School 5. Many detentions and suspensions Added few books to library Added many instructional CDs to classrooms Low student achievement

Let's find (induce, figure out) the connection. Which features that describe the schools throughout the year are, logically, the likely causes of high versus low achievement at the end of the year? Just compare and contrast the schools where high achievement happened at the end of the year versus schools where high achievement **did not** happen at the end of the year.

Can you find features that describe the schools throughout the year when high achievement happened?

Can you find features that that were NOT there throughout the year, and high achievement did NOT happen?

Yes, high quality curriculum materials and instruction are the only school features that were there throughout the year when high achievement happened at the end of the year (Schools 1 and 3), **and they are the only things that were NOT there when high achievement did NOT happen** (Schools 2, 4, 5). You came to (induced) this conclusion----the connection between school features and achievement---through inductive reasoning, by comparing and contrasting examples and finding what goes together and what does not go together.

No doubt, you used the reasoning steps above (comparing and contrasting) without any help from the writer. Your "learning mechanism" (Engelmann and Carnine, 1991)---that is, your brain, language, eyes, etc.---already knows how to perform the steps of inductive reasoning.

- Compare examples with high achievement---1 and 3. Find common features.
- Compare examples with low achievement---2, 4, 5. Find common features.
- Contrast examples with high versus low achievement. Find what is different that seems to MAKE the difference.
- State the connection.

  When___________ happens, then ____________ happens."

  When _________ does not happen, then _________ does not happen."

Look at the figure, on page 35 below, on learning. The learning mechanism uses inductive reasoning to figure out the connections that make up each kind of knowledge—facts, lists, concepts, rules, routines.

The most effective instruction: (1) uses examples that clearly present the features; (2) uses a range of examples because one example isn't enough for the "learning mechanism" to draw a conclusion. [Does school 1 alone tell you what causes high and low achievement? No.]; (**3) contrasts (through juxtaposition) examples (that have high achievement) with nonexamples (that don't have high achievement) so the learner can compare and contrast; (4) uses clear language** so that students can focus on the examples and not get stuck on the teacher's unclear words.

Applying Knowledge Involves Deductive Reasoning.

Above we looked at acquiring new knowledge (of connections) through inductive reasoning. Inductive reasoning (for learning something new) begins with examples and ends with a general idea that connects the examples. *Deductive reasoning is the flip side---applying knowledge to new examples.* **To apply knowledge, the learning mechanism *begins* with a general idea (for example, a rule) and uses the rule to make predictions about a NEW example.** Here is a common form of deductive reasoning. [Remember Aristotle's concept of syllogism on page 16?]

> All plants need water. [Rule---premise---that connects plants and water.]
>
> Cactus is a plant. [A fact.]
>
> Therefore, cactus needs water. [Conclusion. Rule about cactus plants. "I figured that out. It's logical."]

Now let's apply knowledge to a new example, using the rule about achievement.

> In all schools (that we studied) high quality curriculum and instruction were followed by high achievement at the end of the year. [A rule.]

Jones Elementary School (which we didn't study) uses high quality curriculum and instruction. [A fact.]

Therefore, at the end of the year, Jones Elementary School students will probably have ______________ [Predicted fact.]

If you answered "high achievement," it means that you used deductive reasoning. You used knowledge of a general rule about connections between high quality curriculum/instruction and achievement to make a prediction about a new example. Notice that if Jones Elementary actually DOES have high achievement, it makes your rule knowledge even stronger.

Can you see that it's a good idea to have students apply knowledge to new examples? **To do this, you might first have to teach them how to use deductive reasoning in a more conscious way---how to see if new examples FIT what is already known**.

First grader says,

"This new letter---- m ---- has the same shape as the ones I **already** know."

"All the letters with that shape that I already know say mmmm."

"So, this new letter probably also says mmm." [deduction]

Here's another example of deduction, or deductive reasoning. The teacher uses three examples of war to help students to identify common features (kinds of events [e.g., provocations, threats] and the common sequence [e.g., escalating hostility]. This yields **a general theory of war**---a routine for examining wars. What are the events? In what way are they provocations? What is the sequence? Students then USE the general theory of war (the routine for examining war) **to make predictions** about hostilities in the Middle East, or to examine a war that they have not studied.

****

Assignment 3.

Here are data on two **variables**---*classes of stuff whose examples have varying values*. Each row (an example of a general rule/connection) says "When you have this much _______________, you also have that much ________________."

Use inductive reasoning to figure out (induce) a general rule about "this much" in relation to "that much."

State the steps that you used. Did you compare and contrast examples?

	Average Temp	No. of Suicides Monthly/1000
Jan	36	68
Feb	39	80
March	43	86
April	50	102
May	57	105
June	63	100
July	66	100
Aug	65	82
Sept	60	74
Oct	52	70
Nov	43	66
Dec	38	61

Rule = (When, the more, the higher) _________________, (then, the more, the less, the higher) ___________________.

Now make a deduction from the rule. If _________________ goes from ________ to __________, what do you predict will happen to ____________? State the deductive reasoning as a syllogism, using the lines below.

[If it is true that] When __________________, then _____________________.

And if ____________________________.

Then, _____________________________.

Don't you feel smarter already?

What is a Scientific Theory?

A scientific theory is a **set of general statements (propositions) about how things (variables, or concepts) are connected.** ***Concepts*** or **variables** are what the theory is **about**. The ***propositions*** connect the concepts, or variables. Propositions state **relationships** among the variables. A theory explains something (e.g., how students learn general ideas) by stating connections among the variables (concepts, factors) that **PRODUCE** the thing to be explained; e.g., achievement. Here is an example of a theory of learning, in the form of a diagram.

Theory (and Model Diagram) of the Learning Process

From Specific Events	**Students Learn/acquire** →	**General Ideas: Four Forms of Cognitive Knowledge**
Input (Independent) Variables	**Intervening Variables**	**Outcome (Dependent) Variables**
Teacher Presents Examples and Treats Them the *Same* Way (e.g., names, solves, analyzes them). + Teacher presents *Non*examples and Treats Them a *Different* Way **---->**	The Learning Mechanism Performs a set of Logical Operations. It: a. *Examines* examples; *observes* their features b. *Compares and contrasts* examples; *identifies* features that are the same c. *Contrasts* examples (that share some of the same **---->**	Makes generalizations a. Verbal Association (1) simple fact X goes with Y (Name <-> event) (2) list, or verbal chain X goes with Y1-Y5 (New England <-> list of states) b. Concept

(e.g., names, solves, analyzes them).

+

Teacher provides *Assistance* such as Gaining attention, Review, Framing the Task, Modeling the Information, Leading Students Through the Information, Testing/ Checking to Ensure Learning, Correcting Errors, Outcome Assessment.

features and are treated the same way) with nonexamples (that don't have those features and are treated differently).

d. *Identifies the differences* (in the features) between examples and nonexamples, and how they are treated.

e. Makes a *generalization*:

(1) sensory, or basic (All defining features can be seen, heard, felt) E.g., red, on, faster

(2) higher-order (Defining features are spread out and must be synthesized) E.g., Sandstone, justice

c. Rule-relationship, or proposition; that is, statements that tell how concepts/variables are related. E.g., "Frequent practice (one variable) strengthens Retention of knowledge (another variable)."

d. Cognitive routine (Sequences of steps for accomplishing task). E.g., sounding out words, solving math problems, writing essays.

[Adapted from Engelmann, S., and Carnine, D. (1991). *Theory of Instruction*. Eugene, OR: ADI Press.]

A Theory Should Identify the Important Independent/input variables, Intervening variables, Contributing conditions, and Dependent/outcome variables.

Notice that the theory of learning, above, lays out independent/input variables, intervening variables, and dependent/outcome variables. On the left are **input** variables---also called *independent* variables. These are seen as **causes** of (or at least events **prior** to) something else. Prior to what? Students' knowledge---on the right. Students' knowledge is seen as the **outcome** (consequence, effect) of the independent variables. That's why students' knowledge is called "dependent."

Notice the input (independent variables). They include examples and nonexamples that teachers use to communicate a general idea (e.g., concept).

The independent (input) variables also consist of what teachers do to gain attention, model/present information, correct errors, etc. Please identify the rest of the input (independent) variables….

Now look in between the input (independent) and outcome (dependent) variables. There is a set of variables called "**intervening" variables.** In this theory, the intervening variables are the logical operations (what students DO) to GET (induce, figure out) the general ideas (knowledge) revealed by the examples. In other words, the theory states that the examples teachers use (independent variables) **aren't enough** *by themselves* to produce knowledge (the dependent variables). Knowledge also requires the intervening variables of students DOING something with the examples. This says that teachers may have to TEACH students HOW to do this. If students don't know how, then the teacher can present examples in a skillful way, but students will not be able to FIGURE out what the examples say. [This is a huge bone of contention between **instructivists** (who believe that you need to teach both content and HOW to think, and you must carefully design instruction so that it is easy for students to make inductive and deductive generalizations), and **constructivists** (who believe that human beings are already programmed how to think, the teacher should be a guide or facilitator, and you don't need to design instruction carefully because it is good for students to struggle to construct knowledge. Does this need to be a never-ending debate? What does research say? Do students learn to read better when they are taught or when they guess?]

Either way, it is important that this theory lays out the intervening variables. Therefore, a researcher can test WAYS to teach students how to perform the operations, and can then see if THAT (PLUS proper examples) increases learning.

So, the theory really says,

If teachers present a proper set of examples,	→	[And if students perform a set of logical operations	→	Then students will will learn general

and assist students to make sense of the examples,…	with the examples,]….	ideas.
[Independent variables]	[Intervening variables]	[Dependent variable]

Here is another example of a theory---how you catch a cold.

Viral dose →	[If Weak Immune System] →	Likelihood of Cold
[Independent variable]	[Intervening variable]	[Dependent variable]

This theory says, If you receive a sufficient dose of virus, AND your immune system is weak, then you are likely to catch a cold. In other words, the virus isn't enough. You also need a weak immune system.

[By the way, these statements---such as If X, then Y---are propositions. They assert relationships among variables/concepts.]

Here's another.

Tested, effective math materials →	[If teachers communicate clearly] →	Student achievement
[Independent variable]	[Intervening variable]	[Dependent variable]

This theory says (asserts propositions), If you use tested and effective materials, AND if teachers use the materials in a proficient way, THEN students will achieve. In other words, materials alone aren't **enough** for students to achieve. Nor is teaching WITHOUT effective materials. That would be like a surgeon who is proficient but has no tools.

As a consumer, you want researchers to spell out all of the variables in their theory of how things work. What exactly are the dependent variables, the independent variables, and the intervening variables? If researchers leave out the intervening variables, it may mean that they are lazy. It also means that ***they are suggesting that the input variables by themselves have the effect.*** **This is almost NEVER the way things are.** Even catching a cold involves intervening variables. So, if the researcher concludes an article by saying, "X produced the following effects on student learning," **you KNOW that this isn't the whole story**. And you are warned that if you DO X (as the researcher suggests), your students aren't likely to learn, because *there are intervening variables that you don't know about*.

Assignment 4.

Here are some examples. Try to fill in the blanks. What do YOU think important intervening variables might be?

Teacher creates cooperative learning groups	→	[If teacher…..]	→	Accomplishment of group tasks
[Independent variables]		[Intervening variables]		[Dependent variable]

Teacher establishes classroom rules	→	[If …………………..]	→	Students cooperate with rules
[Independent variables]		[Intervening variables]		[Dependent variable]

Imagine that an author tells readers about how she established cooperate learning groups, and about how well the groups accomplished their tasks, but she does NOT tell readers about the **intervening variables** that **made** the cooperate learning groups work. Readers might establish cooperate learning groups in their own classrooms and expect it to work; but it flops. Because the researcher did not tell readers **what else had to be done**.

That is why school reforms (independent variables) often fail (dependent variables)! What features of the school organization (intervening variables) are needed for a reform to work? Can you say "leadership"? Can you say "progress monitoring"? Can you say, "Timely assistance"? What else?

A Theory Should be Consistent with Sound Prior Research and Thinking.

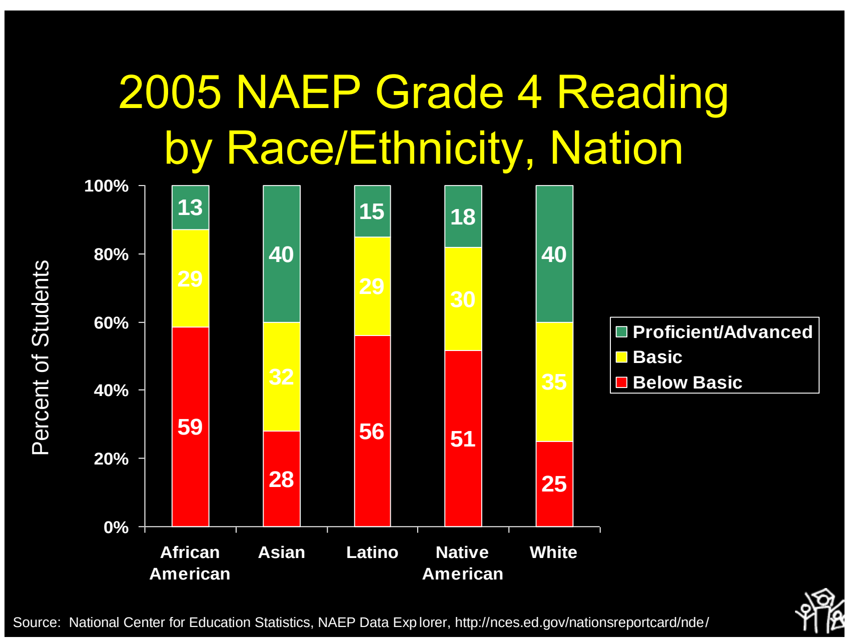
It is important that a theory lay out the dependent variables, the independent variables, and the intervening variables. It is also important **that the theory is derived from or is consistent with sound research and prior thinking.** Otherwise, it's not a theory. **It's a fantasy.** The above theory of learning **is** consistent with a large amount of research. [See references.] However, below is an example of a theory of reading that has been widely used. But this theory isn't based on sound research. **If the theory is false, and if a whole set of methods of teaching and assessment have been derived from the theory, do you think these methods are likely to work?**

> ...I offer this: Reading is a selective process. It involves partial use of available minimal language cues selected from perceptual input on the basis of the reader's expectation. *[Who says?]* As this partial information is processed, tentative decisions are made to be confirmed, rejected, or refined as reading progresses.... *[Does anyone really do that?]* More simply stated, reading is a psycholinguistic guessing game. *[Is it?]* It involves an interaction between thought and language. *[Thinking USES language. There is no thinking WITHOUT language. How can they interact? It's like saying "The saw interacts with sawing."]* Efficient reading does not result from precise perception and identification of all the elements, *[Who says?]* but from skill in selecting the fewest, most productive cues necessary to produce guesses which are right the first time. *[Who says?]* The ability to anticipate that which will be seen, of course, is vital in reading, just as the ability to

anticipate what has not yet been heard is vital in listening. *[An analogy isn't evidence?]* (Goodman, K. (1967). Reading: A psycholinguistic guess game. *Journal of the Reading Specialist*, May, 126-135. pp. 127-8).

This is the theory behind whole language. It asserts that "Efficient reading does ***not*** result from precise perception and identification of all the elements" (that is, readers don't precisely see and identify the LETTERS). Instead, they use other CUES to guess what words say. Goodman seems to admit that he is *making this theory up*. "I offer this…" And he does not cite any research to back up his claims. Yet, for 30 years, tens of thousands of teachers taught millions of children to guess at words, based on this "theory."

Here's the result of poor reading instruction.



What if consumers (teachers) had known that researchers must back up their "theories" with solid research? What if consumers had rejected this theory (and therefore the whole “philosophy” and set of “methods”) because it sounded like a nutty fantasy---that it not only was NOT backed up by research, but did not even make sense? Are YOU guessing what the words say?

What if teachers had seen these before they were swindled?

Telling the difference between baloney and serious claims doc

Telling the difference between baloney and serious claims ppt

Assessing the Quality of Research Plans and Publications ppt

Definitions of Variables

A definition is a statement that tells what a word (a name for a variable, or concept) means, or signifies, or points to.

“Democracy” → (signifies, means, is defined by) → The things over there.

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 variable) means performance that is both accurate and rapid, then to measure fluency you must measure how accurately and rapidly a person does something.

Sadly, 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 search lights that shine on a general area. A conceptual definition of fluency might be:

Fluency is a feature of performance: accuracy and speed.

Here is a conceptual (general) definition of decoding.

Decoding is a routine that involves translating written words into speech, using knowledge of the alphabetic principle (letters have sounds).

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 of the performance to new situations. Likewise, the definition of decoding directs your attention to what **students do** when they read words, and away from things that aren't part of decoding, such as guessing.

Operational definitions. *Conceptual definitions aren't precise enough.* To create actual ways of measuring a 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:

Measuring fluency with grade 1 level connected text.

1. Present sample text.
2. The child reads the text.
3. The observer marks each error—child says different word, adds or omits a sound, says nothing, correctly says word but takes more than 5 seconds.
4. The child reads for one minute. The observer counts the number of errors and subtracts this from the total number of words read.

Likewise, here is a possible operational definition of decoding. First go back and read the conceptual (abstract) definition. Now….

> Decoding is a routine that consists of saying the sounds in a word, from left to right, producing a recognizable word.

Let's line up the pairs of definitions.

Conceptual	**Operational**
Fluency is a feature of performance:	By the end of grade 1,
The combination of accuracy and speed.	the student reads grade level

connected text at the rate of 60 correct words per minute.

Decoding is a routine that involves translating written words into speech, using knowledge of the alphabetic principle (letters have sounds).	Decoding is a routine that consists of saying the sounds in a word, from left to right, producing a recognizable word.

Do you see that the operational definitions **say the same thing** as the conceptual definitions, but are more precise? For instance, the conceptual definition says "translating written words into speech," but the operational definition says "saying the sounds in a word, from left to right…" (a more precise way of saying "translating"). This is precise enough that you can measure it.

****

Assignment 5.

Here are examples of conceptual definitions. *Think of operational definitions for each one*. Remember, the operational definition has to say the say the same thing as the conceptual definition, but it is more precise; it gives examples. Also, operational definitions **depend on the situation**. For example, part of an operational definition of aggression might be hitting, but NOT if you are talking about the sport of boxing!

Conceptual definition	**Operational definition**
Aggression is behavior that is Intended to cause injury	[Aggression on an elementary school playground]
Reading fluency Fluency is a feature of	Second grade reading fluency http://reading.uoregon.edu/flu/flu_benchmarks.php

performance:
The combination of
accuracy and speed.

Proficient instruction is instruction that is carefully designed and delivered and contains features that gain and sustain student attention and participation.

You want a hint? NO!
Let's see how tough you are.

When you evaluate the definitions used in research, ask:

1. **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?
2. **Were conceptual definitions derived from or consistent with scientific research?** For example, reading might be TOO NARROWLY 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). http://reading.uoregon.edu/ **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 based on this NARROW definition will be INVALID.

3. **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.**
4. **Do the definitions consist of words with clear meaning?**
   You saw the theory of learning in the earlier diagram. The words are clear. Examples, compare, contrast, gain attention, etc. Here is another definition of learning. What do you think? Are the words clear?

> "From this perspective, learning is a constructive building process of meaning-making that results in reflective abstractions, producing symbols within a medium." (Fosnot, C.T. (Ed.) (1996). *Constructivism: theory, perspectives, and practice*. New York : Teachers College Press. Fosnot, 1996, p. 27).
>
> "Reflective abstraction is the driving force of learning." (Fosnot, C.T. (Ed.) (1996). *Constructivism : theory, perspectives, and practice*. New York : Teachers College Press.Fosnot, 1996, p. 29).

Do you know what Fosnot is talking about? Do you know what a "constructive building process of meaning-making that results in reflective abstractions, producing symbols within a medium" looks like? If you don't, how could you determine whether Fosnot's data have anything to do with her theory? Why would a person NOT write more clearly? Maybe they can't think straight? Maybe they want to bamboozle you? Here are lines from campaign speeches of a leading national political candidate?

> And he who in **this people** sympathizes with the **poorest** of its citizens, who in this people sees in **every individual** a valuable member of the **whole community**, and who recognizes that this **community** can flourish only when it is formed not of rulers and oppressed but **when all**

according to their capacities fulfill their duty to their *country* and the **community of the people** and are valued accordingly, he who seeks to preserve the native vigor, the strength, and the youthful energy of the millions of **working men**, and who above all is concerned that **our precious possession, our youth**, shouldn't before its time be used up in unhealthy harmful work - he isn't merely a *Democrat*, but he is also *a leader* in the highest sense of that word.

So we have come together on this day **to prove symbolically** that we are **more than a collection of individuals striving one against another**, that none of us is too proud, none of us too high, none is too rich, and none too poor, to stand together before the face of the Lord and of the world in this indissoluble, sworn community. And this **united nation**, we have need of it. When was a leadership at any time faced with a heavier task than our *nation's* leadership?... **What have we? One thing only; we have our people….** On it alone can we count. On it alone can we build. Everything that we have created up to the present we owe solely to its goodness of heart, its capacity, its loyalty, its decency, its industry… And with this people we must and shall succeed in achieving also the tasks that are still to come.

What we want lies clear before us: **not war and not strife**. Just as we have established peace within our own people, so **we want nothing else than peace with the world**.

Sounds great, huh?

You just elected Adolph Hitler.

Propositions

Propositions assert relationships. You can diagram them

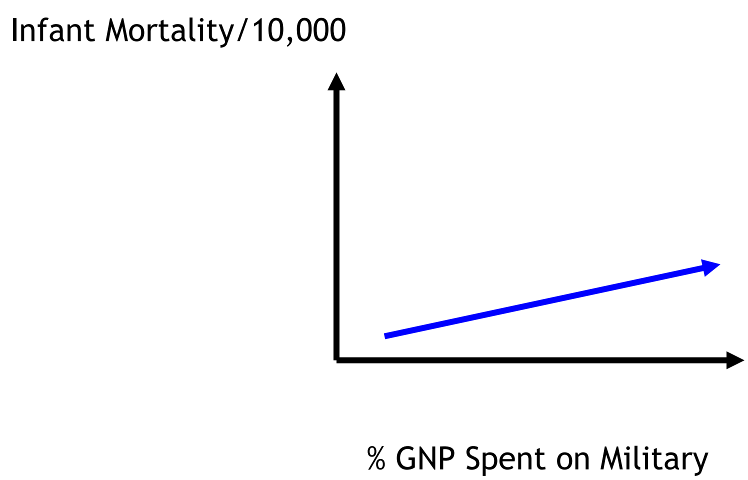
X → Y

or you can say them.

"Whenever X happens, Y happens."

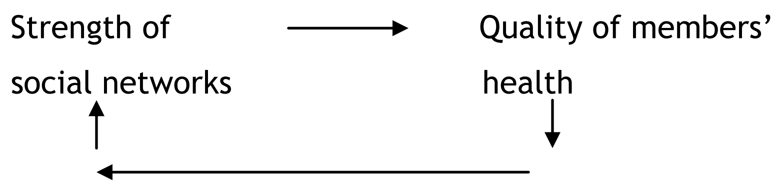
Relationships among what? The answer is, relationships among concepts or variables---classes or families of specific events, things that are in the categories X and Y. Examples of propositions include the following. [We'll define the terms---such as "unilateral"---later.]

1. "**The larger** the percentage of a country's GNP is spent on the military, **the higher** is its rate of infant mortality."
   [This is a **hypothetical/causal/functional proposition** asserting a **direct** relationship (variables change in the SAME direction) that operates in one direction--**unilaterally**. Think about it.]

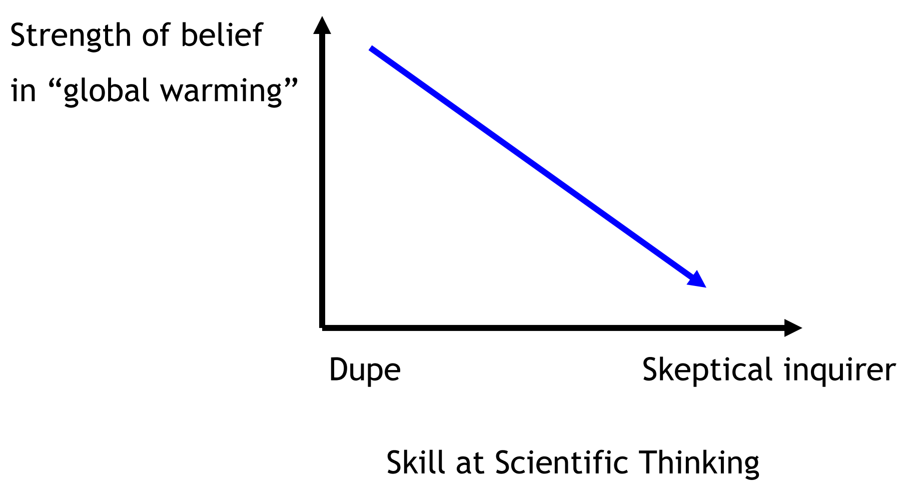


2. "**The greater** the strength of social networks ("strength" operationalized, for example, by the number of people in networks and how often members of a network interact with one another), **the better** is the health of its members."

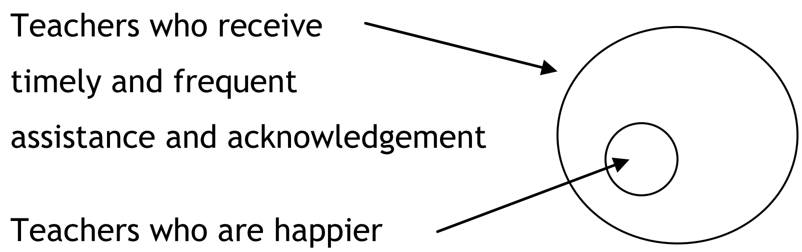
[This is a **hypothetical/causal/functional proposition** asserting a direct relationship that could be bi-directional or reciprocal. Change in one variable affects the other variable. Change in the other variable then effects change in the first---in a circle. Strong social networks sustain health. Health makes it possible to participate in social networks.]



3. "**The stronger** the social integration in a community (operationalized, for instance, by the percentage of eligible voters who vote, the percentage of families that attend some kind of church services on a regular basis, the percentage of eligible or relevant persons who attend PTA meetings, the average number of neighbors whom persons can name), **the lower** is the rate of suicide, alcoholism, and juvenile crime."
[This is a **hypothetical/causal/functional proposition** asserting an **indirect or inverse** relationship (variables change in opposite directions) that might be reciprocal. One variable increases and the other variable decreases.]



4. Teachers who receive timely assistance and frequent acknowledgement of proficient teaching rate themselves as happier on the job.
[This could be a **categorical proposition**; it asserts that items in one category (teachers who are happier on the job) are **included** in another category (teachers who receive timely and frequent assistance and acknowledgement).



However, it might also be considered a hypothetical/causal/functional proposition; timely assistance and frequent acknowledgement increase happiness.]

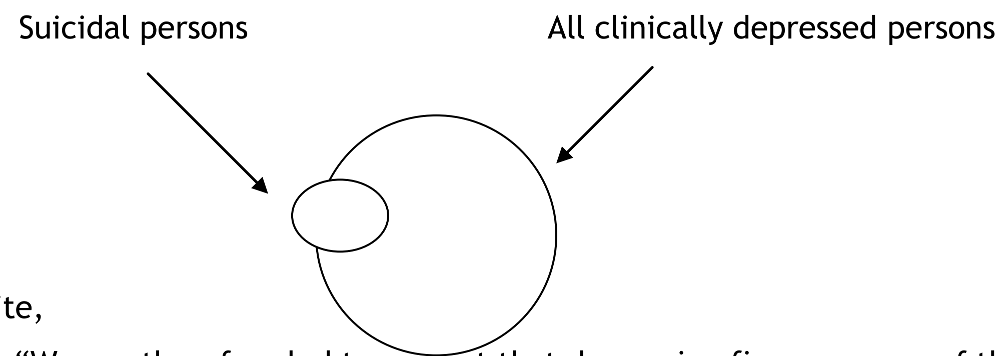
Timely assistance and frequent acknowledgement → Happiness

Some theorists and researchers are easy to read because they link propositions in a logical way--one proposition leads to the next. The sequence is like a logical argument—a **routine** (one form of knowledge) for explaining something. However, many writers:

1. Scatter propositions around, and so the reader can only speculate about what the argument (the flow of logic) is. "Huh? What's she saying?"

2. Fail to state propositions in good propositional form. For example, instead of a straightforward statement, such as,

"Most suicidal persons are clinically depressed." [Categorical proposition]



they write,

> "We are therefore led to suspect that depression figures as one of the most important features in the etiology of suicide." [Just say it, will ya?]

3. Fail to state definitions in proper definitional form; e.g., "By 'aggression' is meant behavior (**genus**) that is intended to injure a living thing (**difference**)." Poor definitions leave the reader guessing what the writer means.

4. Contradict themselves, change definitions, or use vague definitions. The result is endless dispute about what the writer "really said." Or the writer is considered profound because no one knows what he or she is talking about.

Categorical and hypothetical (causal/functional) relationships

Propositions generally assert two kinds of relationships: categorical and hypothetical.

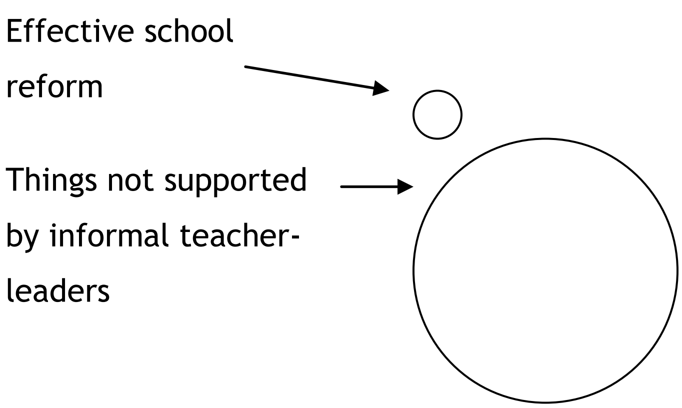
Categorical propositions. Categorical relationships (one thing is part of, not part of, or partly part of another thing) are asserted by categorical propositions. Following are examples. You can use Venn diagrams to illustrate inclusion and exclusion.

1. "**All** proficient readers know how to sound out unfamiliar words." [This categorical proposition asserts that one category is completely **within** another category.]

   "**No** students who guess what words say read fluently and with high comprehension." [This categorical proposition asserts that one category is completely **outside** another category.]

2. "Some teachers only select curriculum materials that have been tested with level 3 evaluation research." [This categorical proposition asserts that **part** of one category is **within** another category.]

3. "School reform isn't effective when informal school leaders (e.g., teachers) don't support it." [This proposition asserts that none of one category is in the other category.]



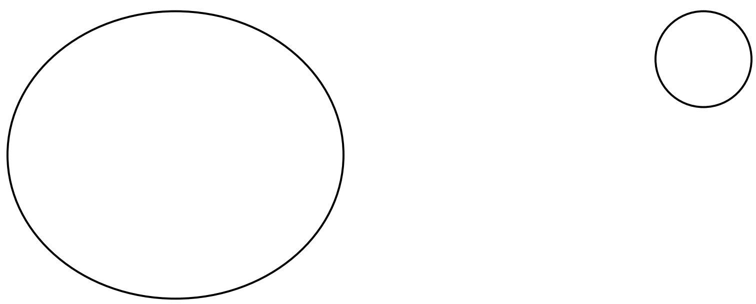
In summary, categorical propositions assert that all (or part) of one class, concept, or variable is included in or is excluded from another class, concept, or variable.

Assignment 6.
Write and diagram categorical propositions regarding the following sets of two variables: (1) things fostered by all skilled teachers and achievement in students; (2) successful school reform efforts and social systems in which

members **don't** have a shared mission; (3) adults with antisocial personalities and children who receiv**ed** harsh discipline (all or some?); (4) effective leaders and persons who are trusted.

Here are some hints. Take number 1. Which is the larger category---achievement in students or things fostered by all skilled teachers? Which category has more stuff in it? Do skilled teachers foster achievement? Yes. Is that the ONLY thing they foster? No. Do they foster other things, too? Yes. So,

Things fostered by all skilled teachers. Student achievement.



Where does achievement go? Outside, inside, or partially inside things fostered by all skilled teachers?

The proposition would be: "All _______________________ foster __________________."

Causal/functional propositions. Causal/functional relationships are asserted by hypothetical or causal propositions. One thing influences (causation) or changes along with (correlation) another thing. Below are several examples.

1. "The more stressors that bear upon people during a year, the more illnesses they will have during that year."

This causal/functional hypothesis or hypothetical proposition asserts a **direct relationship** between stressors (independent variable) and illness (dependent variable); i.e., as one variable changes in one direction (up or down) the other variable changes in the same direction. Either both variables increase or both decrease.

2. "The more interpersonal support persons have for their moral principles (independent variable), the less likely they are to obey orders which prescribe what they consider immoral acts (dependent variable)."

The above causal/functional hypothesis or hypothetical proposition asserts an **inverse (or indirect)** relationship between interpersonal support and obedience. As one variable changes in one direction (up or down), the other variable changes in an opposite direction.

Hypothetical (or causal/functional) propositions assert that the existence of or a change in a dependent variable (the consequent or alleged effect) is preceded, predicted, determined, dependent or **contingent** upon the existence of or a change in an independent variable (the antecedent or alleged cause). However, **there are several degrees and types of dependence or contingency**. For example, independent variables may be seen as necessary conditions, sufficient conditions, intervening variables, and contributing conditions.

1. necessary condition. The existence of or a change in the dependent variable **requires** the existence of or a change in the independent variables. For instance:

"**If and only if** there are shared feelings of exploitation among subjects, will subjects mount resistance against rulers whom they perceive to be exploiting them."

***2. sufficient condition*.** The independent variable isn't asserted to be a necessary condition; it is assumed that **other** independent variables **also** can have the asserted effect on the dependent variable. However, the independent variable is asserted to be sufficient (enough by itself) to effect a change in the dependent variable. For example:

"**Whenever** there are shared feelings of exploitation among subjects, they will mount resistance against the rulers whom they perceive to be exploiting them."

Generally, no one factor is likely to be sufficient. Instead, a set of necessary conditions (e.g., shared feelings of exploitation **plus** an opposition ideology **plus** opposition leaders **plus** opportunities to mount resistance) is usually asserted to make up a sufficient condition. This **set** of independent variables may operate in a sequence or in a configuration, as shown.

Independent variables in a **sequence**:

If V, then W; if W then X; if X, then Y; if Y, then Z (final dependent variable)

Or,

Independent variables in a configuration:

```
V <------> W <-----> Z
^
|
v
X -------> Y
```

***3. intervening variable*.** Some variables are neither necessary nor sufficient. Rather, they **stand between** main independent variable(s) and the dependent variable(s). W, X, and Y, above, are intervening variables--i.e., intervening between the more distant effects of V on Z. For example, it is generally true that the larger the dose of cold virus, the greater the likelihood that people will catch a cold. However, the relationship between viral dose (independent variable) and the probability of catching cold (dependent variable) is influenced by a third variable--namely, the strength of the immune system. In other words, viruses produce colds (they are necessary conditions) but generally **only if** the immune system is weak enough. In a causal model of these relationships, the strength of the immune system is a **gatekeeper** standing between viruses and colds, as shown.

Viral dose -------> [If Weak Immune System] --> Likelihood of Cold
Main Independent --> Intervening Variable -----> Main Dependent Variable
Variable

It is seldom easy to determine if a variable is an intervening variable. We must compare situations in which the alleged independent variable exists, but the possible intervening variable **sometimes exists and sometimes does not exist**. For example, participants in an experiment get different doses of cold virus. Some receive a large dose; some a moderate dose; and some a small dose. Seventy-five percent of those receiving a large dose shortly caught a cold; half receiving a moderate dose caught a cold; and only ten percent receiving a small dose caught a cold. In other words, **the larger the dose of virus, the higher the probability of a cold (empirical generalization).** But suppose we also measured the strength of each person's immune system. Let us **statistically remove** from the sample (take out of the data) all persons with a strong immune system, and then **re-analyze the data only with persons having a *weak* immune system.** Now we find that ninety-five percent of the people receiving a large dose got a cold (it was only seventy-five percent when those with a strong immune system were in the high-dose group); seventy percent of those receiving a moderate dose got a cold (it was fifty percent before those with a strong immune system were taken out of the sample); and thirty percent of those receiving a small dose of virus got a cold (it was only ten percent when persons with a strong immune system were in the sample).

The findings show that **the strength of the immune system makes a difference in whether people get a cold.** By itself a weak immune system isn't sufficient to cause a cold; one still needs a dose of virus. Nor is a weak immune system a necessary condition for catching a cold, because some people with a strong immune system still do catch a cold. (It could be that even strong immune systems are overwhelmed by certain strains of cold virus.) Therefore, the correct empirical generalization seems to be this--The larger the dose of cold viruses (and to the extent that the immune system is weak), the greater the likelihood of catching a cold.

4. contributing condition. A contributing condition affects the amount, type, or speed of change that can be effected by the main independent and intervening variables. For instance, whether people get sick depends upon the size of the viral dose (the main independent variable) and the strength of the immune system (intervening variable). But **how long** people remain sick may have little to do with dose and immune system. Rather, it may be a function of personality traits (such as healthy-mindedness), diet and rest during the illness, pressure to return to work, or rewards for acting sick.

Here is another example of a contributing condition. When subjects in an authoritarian social system collectively realize that the costs of submission far outweigh the rewards they receive in exchange, the likelihood of resistance to rulers increases. But **what kind** of resistance will subjects mount? Will it be private grumbling, peaceful demonstrations, work stoppages, or violence? The kind of resistance may be a function of the amount of violence rulers have used against subjects. Thus, rulers' use of violence may **contribute** to the form of resistance, but it may not affect the **likelihood of** resistance. How do we determine the causal **function** of independent variables (i.e., as necessary, sufficient, intervening, or contributing)? The answer is that we construct a tentative (hypothetical) causal model, and conduct research to test the model.

Direction of causal/functional relationships. Causal/functional propositions generally assert a causal "flow" or "path" among the variables. These paths are as follows.

1. Unilateral. Unilateral relationships are **in one direction only**. That is, change in an independent variable (necessary condition, sufficient condition, intervening variable, or contributing condition) effects a change in the dependent variable, but the change in the dependent variable does **not** then affect the independent variable. For example, something about social class (degree of frustration? models of violence?) affects the rate of homicide in each social class, but *the rate of homicide does not cause social class*.

2. Bilateral or reciprocal. A bilateral relationship is **two-way**. Change in X engenders change in Y; the change in Y then effects a further change in X. This reciprocal (back-and-forth) relationship is called a **feedback loop**. Feedback loops are of several kinds. One kind is a **positive** feedback loop. In a positive feedback loop, each increase (or decrease) in one set of variables effects a **further** increase (or a further decrease) in the other set of variables. That is, each set either **amplifies or dampens** the other set in the same direction. For example, in a "heated argument," the behavior of one person fosters an increase in the "heat" of the other person's behavior, which fosters **even more** "heat" in the first person's behavior, which produces still more "heat" in the other's behavior, until some limit is reached. Or, as one person withdraws in a relationship, the other person may withdraw some, which results in the first person withdrawing more than before, which results in the other person withdrawing even more than before, until a **limit** is reached (separation).

Another kind of reciprocal influencing is a **negative feedback loop.** In a negative feedback loop, change in one set of variables effects an increase, say, in the other set of variables. The increase in the second set then results in a **dampening** or a decrease in the level of the first set. For instance, the heat that comes from a furnace raises the temperature of the room until the temperature is high enough to shut off the furnace. Or, an increase in the rate of crime in a city produces an increase in the number of police in the city, which results in a decrease in the rate of crime. Of course, the decrease in the rate of crime may result in a decrease in the number of police, which then results in another increase in the rate of crime, and another cycle begins. This would be an example of **oscillation**.

3. Dialectical. A dialectical relationship involves reciprocal influencing, but with one more feature. As each set of variables influences the other set, the **quantitative** changes eventually yield a change in the **quality, type, or state of each variable, and also perhaps in the nature of the relationship.** For instance, at 33 degrees Fahrenheit, if one more degree of heat is lost, the water becomes ice. Or, if parents accidentally reward their young children for throwing tantrums and hitting, the children will perform these behaviors more often. The parents then try harder to stop the problematic behaviors in ways that, again, reward these behaviors. At some point, **quantitative changes in the children's behaviors result in a qualitative shift** in the way the children are perceived. They are no longer seen as normal children who perform problematic behavior too often; they are seen as children with a conduct disorder. At the same time, the parents no longer see themselves as regular parents, but as guards or victims. Finally, as the nature of each person's participation in the relationship changes, the nature of the relationship itself changes; e.g., from sweet children and loving parents (a complementary relationship) to an adversarial relationship (a symmetrical relationship).

Think of dialectical changes in a school (e.g., between leadership, instruction, and student achievement) that eventually yield a different KIND of school.

4. Configurations, networks, and ecological systems. Social systems contain many interrelationships among many variables (features). To make matters more complicated, many interrelationships are reciprocal and/or dialectical. Indeed, a system may be so complex that it is hard to determine which variables and relationships are more important in fostering certain outcomes. In fact, if we study some relationships in isolation from the system in which they ordinarily occur, the results may not reflect how things usually are but only how they appear in a contrived situation.

Proximity. Some causal/functional relationships are **"proximal."** That is, there is little time lag or few intervening variables between the main independent variable and the main dependent variable. Other causal/functional relationships are **"remote" (distal)**. Sometimes, remote causes are considered **predisposing** factors and proximal causes are considered **precipitating** factors. However, these terms are vague with respect to the degree of dependence. For instance, if early childhood experiences are considered remote causes of adult emotional difficulties, are those childhood experiences necessary conditions, sufficient conditions, contributing conditions?

Assignment 7. Concepts, Definitions, and Propositions

Following are excerpts that contain definitions and propositions. Find these and then state them in proper propositional and definitional form. Note that many propositions and definitions are implicit; e.g., the logical flow from proposition 1 to proposition 2 requires another (unstated) relationship or a definition. Also identify if there is reason to believe that the relationships are direct vs. inverse; uni-lateral vs. bi-lateral; involve necessary, sufficient, or intervening variables.

Yes, doing this WILL hurt your head. BUT you'll be so much sharper!!!

1. ...a state is a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory. [Max Weber. "Politics as a vocation." 1918*] Is he asserting causation or is he defining a concept?*

2. No living being can be happy or even exist unless his needs are sufficiently proportioned to his means. [Emile Durkheim, *Suicide*. 1897]

3. If the state is to exist, the dominated must obey the authority claimed by the powers that be. [Max Weber. "Politics as a vocation." 1918] *Is he asserting that something is a necessary condition for another thing?*

4. ...the term suicide is applied to all cases of death resulting directly or indirectly from a positive or negative act of the victim himself, which he knows will produce this result. An attempt is an act thus defined but falling short of actual death.   [Emile Durkheim, *Suicide*. 1897] *Causation or definition?*

5. *If therefore industrial or financial crises increase suicide*, this isn't because they cause poverty, since crises of prosperity have the same result; *it is because they are crises*, that is, *disturbances of the collective order*.   [Emile Durkheim, *Suicide*. 1897] *Both causal and categorical propositions here*.

6. Where the State is the only environment in which men can live communal lives, they inevitably lose contact, become detached, and thus society disintegrates. [Emile Durkheim. *The Division of Labor in Society*. 1893]
Do you see a causal chain?

7. There is the authority of the extraordinary and personal gift of grace (charisma), the absolutely personal devotion and personal confidence in revelation, heroism, or other qualities of individual leadership. This is charismatic domination... [Max Weber. "Politics as a vocation." 1918]

8. Quoting Esquirol (with whom Durkheim disagrees):

"Suicide shows all the characteristics of mental alienation." (p. 58)

"Mental alienation" means insanity. Which is the more inclusive category?

9. Average of Suicides per Million Inhabitants [Do you see a range of orthodoxy here?]

Average of Suicides per Million Inhabitants	
190	Protestant States
96	Mixed States (Protestant and Catholic)
58	Catholic States
40	Greek Catholic States

So, you could say "The more….., the higher the…."

10.

Provs with Cath Minor (<50%)	Suicides/ Million Inhab	Provs with Cath Major (50-90%)	Suicides/ Million Inhab	Provs with More Than 90% Cath	Suicides/ Million Inhab
Rhenish	167	Low. Franc	157	Upp. Palatin.	64
C. Fracon.	207	Swabia	118	Upp. Bavaria	114
Upp. Franc	204			Low. Bavaria	19
	Ave. 192	Ave. 135		Ave. 75	

What *empirical generalizations* can we draw from the above table?

"(S)uicides are found to be in ___________ proportion to _____

and in _______proportion to________________________" (p. 153)

11. "(W)hen religious intolerance is very pronounced, it often produces an opposite effect. Instead of exciting the dissenters to respect opinion more, it accustoms them to disregard it." (p. 156)

12. "(A) religious society cannot exist without a collective credo." (p. 159)

13. "(T)he more extensive the credo the more unified and strong is the society." (p. 159)

14. "(T)he greater concessions a confessional group makes to individual judgment, the less it dominates lives, the less its cohesion and vitality." (p. 159) Causal sequence?

15. "Man seeks to learn and man kills himself because of the loss of cohesion in his religious society; he does not kill himself because of his learning." (p. 169)

16. "(T)he desire for knowledge wakens because religion becomes disorganized." (p. 169)

17. "(T)he density of a group *[rate of interaction]* cannot sink without its vitality diminishing." (p. 201)

18. "Excessive individualism…frees man's inclination to do away with himself from a protective obstacle… (p. 210)

19. "...they are crises, that is, disturbances of the collective order." (p. 246)

20. "...more depressed and anxious pregnant teenagers, who perceive their social relationships to be less satisfying, and who have less knowledge of child development, have more negative expectations for their infants." J.M. Contreras et al. (1995). *Journal of Applied Developmental Psychology, 16,* 283-295. *Do you see the intervening variables?*

21. High mother support was associated with more secure infant attachment only for those adolescents living with partners." S.J. Spieker (1994). *Developmental Psychology, 30,* 1, 102-111. *Do you see an intervening variable?*

Measurement

Once a researcher has defined variables conceptually and operationally, the researcher can begin to select or to develop methods of measurement. There are several guidelines that should be followed.

1. Measures should be consistent with the definitions of the variables.

For example, if fluency with math problems is one outcome (dependent) variable, the researcher needs to measure accuracy and speed with which students solve math problems. A measure might be the rate of correct and incorrect problems solved per minute. Likewise, if one input (independent) variable is the **faithfulness** with which teachers follow a written instructional protocol, then the researcher cannot just measure (describe) HOW teachers teach, but must measure how teachers teach **in relation to** the written protocol. The researcher would have to describe the teaching methods in the protocol AND how the teacher USES those methods.

2. Measurement should be direct.

When persons have a lung infection, they often have a fever with it. What would you want your physician to measure, to see if you are getting well: the amount of infection in your lungs, or your temperature? Temperature is an INDIRECT measure of lung infection. And it may NOT be valid. Your fever may be gone but you still have an infection. Likewise, if reading proficiency is an outcome variable, then reading proficiency (e.g., accuracy and speed of decoding, comprehension of text) is what you should measure. How much students enjoy reading, or how much they read outside of school are INDIRECT measures of reading proficiency. Students who read well are likely to enjoy reading and to read more. But these measures may not be valid; some students read a lot, but not well.

3. The researcher should measure at the proper *level* or *scale* of measurement.

Consider the variable, color. There are four "scales" or "levels" for measuring it.

a. You could simply take each color sample and **name** it---say the category it is in. This is called "nominal" level measurement. Think of "name."

b. You could **rank** each color sample from lighter to darker.

Darkest red

Dark red

Medium red

Light red

Lightest red

This is called "ordinal" level measurement. Think of order.

c. You could use a scale of **equal intervals**.

"How much red would you say is in this fabric?"

1 2 3 4

|____|____|____|____|

very little mostly

This is called “interval level” measurement.

d. You could use an instrument that **measures exactly how much white** is in each color sample. The instrument gives you a **number**. This number is a measure of brightness. This is called “ratio-level” measurement. One sample may have 25 white units. Another may have 50 white units. The first one has half the amount of white as the second. The ratio is 1 to 2. Ratio level.

Let’s look at each level or scale in more detail. Here are some useful websites.

http://web.uccs.edu/lbecker/SPSS/scalemeas.htm#3

http://www.math.sfu.ca/~cschwarz/Stat-301/Handouts/node5.html

http://allpsych.com/researchmethods/measurementscales.html

http://en.wikipedia.org/wiki/Level_of_measurement

http://www.kimberlyswygert.com/archives/002750.html

Again, there are four levels of measurement: nominal, ordinal, interval, and ratio. **Each next level provides more precise information than the others.**

Nominal level. The lowest level of measurement. Nominal level or nominal scale measurement implies **qualitative** (type) not quantitative (amount) differences. It refers to kinds or types of things. *Nominal measurement consists of naming or putting the things measured into categories*. For example, you could categorize students into two groups: students who receive free and reduced lunch and students who don’t receive free and reduced lunch. Other examples of nominal measurement include marital status (married, divorced, separated, single), occupation, and ethnic identity.

If you are measuring some variable (e.g., error correction) on a nominal scale, you would *simply put each instance of error correction in one of several types that you had already identified*. For instance, one type might be modeling the correct answer. Another type might be explaining why the student made an error. The third type might be calling on another student to demonstrate the correct answer. After you have collected the data (put all instances of error correction in the proper categories), you would *summarize the data simply by counting the number of instances in each category*.

Data on how teacher corrected math errors during one lesson

Modeled correct answer and then tested……………..12

Explained why student made error………………………..20

Called on another student to come to the …………….8

board and show the correct way.

With NOMINAL data, you can

(1) Figure out how many instances are in each category.

(1) Figure out the percentage of the total that is in each category.

Model and test = 12/40 = 30%

Explain = 20/40 = 50%

Call on another = 8/40 = 20%

(3) Figure out the most frequent category. Explaining = 20. The most frequent category is the **mode**, or the **modal** category.

Please restate the three ways that you can summarize NOMINAL data?

Ordinal level. An ordinal-scale or ordinal-level measure implies a **rank order of degrees or amounts of something, but *not equal intervals* between the degrees or ranks**. Probably most opinions---attitudes, perceptions, and feelings---are in reality 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 aren't* ***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. ***This is why you cannot give a number to each rank, and then <u>add up</u> the rank scores (2, 3, 3, 2, 2, 2, 1, 1, 3, 3, 2, 2) and then divide by the number of scores (12) and find the <u>average rank</u>!*** **Because the distances between the ranks aren't equal.** <u>The NUMBER of a rank isn't a numerical VALUE.</u> It isn'thing more than the NAME of a rank. So, if you measure things by giving their rank order (e.g., you assign each student the rank Proficient/advanced; Basic; or Below basic), you **properly** summarize the data by simply

(1) Figuring out **how many students are in each rank** and then perhaps figuring out the **percentage** of the total number that is each rank. For example, there are 12 students.

Proficient/advanced = 4 = 33%

Basic = 6 = 50%

Below basic = 2= 17%

If you then use a better reading program, you hope that the DISTRIBUTION of rankings changes to, for example,

Proficient/advanced = 4 = 33%

Basic = 8 = 67%

Below basic = 0

(2) Figuring out the most frequent rank, or the **mode**---which, above, is "Basic."

(3) Figuring out the rank that is in the middle---about 50% of scores are above and below it. Here are the data from above. 1 = Below basic; 2 = Basic; 3 = Proficient/advanced.

1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3

The middle of the distribution is 2. **This is called the median.**

Here is another distribution. Income for nine persons.

$100,000

$60,000

$60,000

$20,000

$20,000

$20,000

$10,000

$8,000

$8,000

What is the middle score—about half are above and half are below it? ***$20,000***

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 (that is, well-organized, lots of instructions, easy to use) your new math materials are."

1 2 3 4

|____|____|____|____|

Less friendly More friendly

Or, "How much do you agree with the following statement? "Our school provides timely and adequate supervision and assistance?"

1. Strongly agree.

2. Agree
3. Disagree
4. Strongly disagree.

When it is assumed that the intervals are equal, it is okay to summarize scores by calculated the **mean, or average.** *You add the scores and divide by the number of scores.*

For instance, here are the scores of 10 persons on the above question.

3 persons gave a rating of 3, or 3 x 3 = 9.

4 persons gave a rating of 2, or 2 x 4 = 8

3 persons gave a score of 1, or 1 x 3 = 3

Total score = 9 + 8 + 3 = 20. 20 total divided by 10 scores = 2. The average or mean score is 2.

Ratio level. Ratio-level or ratio-scale measurement is real numbers. There can be true zero (e.g., zero episodes of aggression occurred; zero income). In addition, there are **equal intervals between quantities**; e.g., the difference between 0 and 1, 1 and 2, etc., is 1.

Ratio level measurement is the most precise. It provides information on 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 **count instances** of identified variables (e.g., correct answers).

With ratio-level measures you can do many operations to summarize data. Here are data on reading fluency.

Billy = reading 100 correct words per minute
Sam = reading 90 correct words per minute
Slim = reading 90 correct words per minute

Darren = reading 110 correct words per minute

Nancy = 80 correct words per minute

Terri = 90 correct words per minute

Tim = 95 correct words per minute

(1) Figure out the mode, or most frequent score. 90.

(2) Figure out the median, or the middle score.

80, 90, 90, 90, 95, 100, 110 = 90 (3 scores are above and 3 are below 90)

(3) Figure out the mean, or average.

80 + 90 + 90 + 90 + 95 + 100 + 110 = 655, divided by 7 scores = 93 mean or average score

4. Figure out percentages. For example, the mean fluency when the teacher used *Phud Phonics* was 93 correct words per minute, and the mean fluency **after** the teacher used a new reading program (*Fluent Phonics*) for three months rose to 100 correct words per minute. What is the percentage increase?

From 93 to 100 = increase of 7

What percentage of 93 is 7?

7/93 = approximately 8%

Going from a mean of 93 to a mean of 100 is an increase of about 8%.

A few cautionary comments

1. **You *can* use a lower-level scale for measuring a variable that could be measured on a higher level, but you lose information.** For example, you can measure fluency on a nominal scale by categorizing each student as either “Rapid,” “Moderately fast,” or “Slow.” But this means that several students that are in the same category could actually have different EXACT

fluency rates. You might treat these students the same (e.g., put them in the same reading groups based on their nominal category), when they are actually different. It also means that you don't know EXACTLY how many words students read correctly per minute. Therefore, **it is best to use the highest (more precise) level of measurement that you can.**

2. **However, you CANNOT (!!!) use a higher-level scale to measure a variable that is *really* on a lower scale.** For example, the three different methods of error correction (above) are just categories. The categories don't imply differences in the amount or quantity of anything. Therefore, you cannot give each category of error correction a number…..

    Model correct answer is 1

    Explain error is 2

    Another student demonstrates is 3.

   And then add up the number of 1's, 2's, and 3's…..

    Model and test = 12 12 x 1 = 12

    Explain = 20 20 x 2 = 40

    Call on another = 8 8 x 3 = 24

   And then figure out the mean…..

    12 + 40 + 24 = 76 76 divided by 40 scores = 1.9 = average or mean error correction.

   This makes no sense at all. The different kinds of error correction aren't WORTH any points. Explaining (a 3) isn't worth 3 times modeling (a 1). These numbers are no more than names.

4. **When possible, the researcher should have several measures of the same variables.**

This is called "triangulation." The idea is, if different measures say much the same thing, you can have greater confidence in the validity of the finding. For instance, a researcher might give students mastery tests every 10 lessons of a math program. The tests are based on curriculum materials that were covered. At the end of the semester, the researcher also gives students a standardized test on math. If the curriculum based measures and the standardized test (that has different kinds of items on it) both say that students have learned the material, then you can have more confidence in the findings than if you had only one measure.

Here are resources on standardized tests.

http://www.ncrel.org/sdrs/areas/issues/students/earlycld/ea5lk3.htm
http://nces.ed.gov/nationsreportcard/about/
http://en.wikipedia.org/wiki/Standardized_testing
http://www.sizes.com/society/test_scores.htm

5. Researchers should assess and report the reliability of measurement.

Observers and testers should be trained ahead of time to follow a testing or observing protocol---steps on exactly what to do. They should be observed testing or observing, and coached to use the protocol faithfully. Scores of the SAME observer or tester scoring the same thing several times should be compared to see how much the two sets of scores agree. This is called **intra-observer** (within the same observer) **reliability**. Also, different observers or testers scoring the same thing should be compared---again to see how closely they agree. This is called **inter-observer** (between observer) **reliability**. If reliability (agreement) is below 90%, then either observers and testers need more training, or the definitions of variables need to be clearer (maybe they disagree because the definitions are vague), and the protocols need to be made easier or clearer. Researchers should describe how they trained observers and testers, and how they assessed reliability. **If not, the consumer has no way to tell if the scores are valid.**

Analyzing Statistical Data

Please examine the entries in "Vocabulary" at the end of this document for mean, median, mode, plot on a graph, relationship, and trend.

Let's say authors are reporting **survey research** of schools that used one of two kinds of math programs. Program A (there were several versions) taught all of the elementary math concepts and operations **before** it had students apply these skills to word problems. Program B (there were several versions) focused on word problems, and taught students the relevant math concepts and operations **at the same time**. The authors believe that Program A will yield higher achievement. So, they divide the schools in the district into schools that use Program A vs. Program B, and they also collect information using district official statistics on the percentage of students who pass end of grade tests (as an outcome measure of achievement.) The authors report, "In general, students who received Program A achieved significantly more than students who received Program B."

Will you use Program A? Will you avoid Program B?

The authors don't tell you what it means that "students who received Program A achieved significantly more than students who received Program B." They are leaving out essential statistical information.

Summary statistics

For each class in each school that used Program A and Program B, you want to know:

1. **The average score---the mean**. The sum of all of the scores in a class divided by the number of scores.

$$\frac{65+ 69 + 70 + 75 + 78 + 80 + 80 + 87+ 93 + 93+ 96}{11} = \frac{886}{11} = 80.5$$

2. Notice that the mean is 80.5, but scores range from 65 to 96. **Range is another statistic to present.** Shouldn't consumers know that a program can produce a WIDE range of scores? Wouldn't you want to know if a medication produced a wide range of effects?

3. **The most frequent score.** This is the **mode**, or **modal** score. What is the modal score?.... 80. Again think of medicine. Can you imagine asking your physician, "What is the **most likely** outcome?" Of course.

4. **The middle score.** This is the **median**. This is an important statistic. It tells you which score is about half way in the **distribution** (spread) of scores.

What is the median score from the above distribution? **80.** Why is it important? Well, imagine that five students scored in the 90s. These scores make the mean or average pretty high. If the mean were the **only** statistic you had, you might think that the class **as a whole** did well. But what if the middle score was 80? Half of the class got lower scores than 80. So, **the median tells you not to be fooled by a high mean that is really the result of a few very high scores, or vice versa.**

5. Percentages or ratios vs. whole numbers. Don't be satisfied if a researcher reports percentages or ratios but not the whole numbers—raw numbers. One researcher reported that students who received a certain pre-school program (vs. a different pre-school program) were **twice as likely**---two decades later--to have been arrested for felonies. **Many readers were completely fooled by this statistic.**

> "Boy, I'm never going to use THAT program. It makes kids twice as likely to become criminals!"

Sure, that's how it looks if you only report percentages and ratios (twice as likely). But what if you found out that after 20 years there were **only three persons left in the samples** for each pre-school program? And what if "twice as likely" means that in one pre-school sample, ONE adult had been arrested for felony, and in the other pre-school sample, TWO adults had been arrested. In other words, percentage-wise, the difference is 200 percent. But in terms of whole numbers or raw numbers, we are talking about ONE person. Do you think THAT is significant? Could it just as easily be a difference of ONE arrest in the OTHER sample? Of course. So, if authors don't report the raw numbers, you have NO idea if the percentages and ratios are meaningful. **<u>200 percent more of WHAT? One!?</u>**

Statistical Significance

In the survey, above, the researchers collected data on student achievement when students used one of two kinds of math programs. They report that "students who received Program A achieved significantly more than students who received Program B." We wondered what that meant. The researchers told us PART of what that meant by giving us summary statistics for each class: the range of scores, the mean score, the modal score, and the median score. The researchers' claim, remember, is that the mean, median, and modal scores for students in Program A are significantly higher than the mean, median, and modal scores of students in Program B. But what does "significantly higher" mean? Significance means two things: practical and statistical.

Practical significance. You join a program to lose some weight.

> "I can stand to lose a few pounds. I have to walk sideways through the doorway."

So, you join Whale Watchers. You pay 100 dollars a month for advice, feedback, encouragement, and menus. At the end of one year, you have lost 10 pounds!

Wow!

1200 dollars.

10 pounds.

Would you say that the result is of **practical significance?**

Can you walk straight through a doorway?

Can you fit into your swim suit?

Can you see your feet?

NO?

So, 10 pounds isn't of practical significance.

Statistical significance. But what if almost everyone in Whale Watchers (thousands of persons) lost from between 5 and 15 pounds? What are the odds of that, if Whale Watchers did not work? What are the odds that so many persons losing weight--even if it is only a little weight---is a fluke, random, chance? ***That is what statistical significance is about.*** If you have large samples, even small but consistent differences between the samples on some outcome measure are probably statistically significant---NOT likely to be the result of chance. [However, small differences may not be practically significant. Would you change an entire reading program just because program C produces on average 2 points higher achievement?]

At the same time, **with small samples, it takes larger differences for the differences to be statistically significant.** Imagine two weight loss programs. Whale Watchers and Pie Anonymous. Imagine that there are five persons in each group. At the end of the year, the mean weight loss in the five Whale Watcher clients was 6 pounds, and the mean weight loss in the five Pie Anonymous clients was 7 pounds, or 8 pounds, or even 9 pounds? Do you think those differences COULDN'T EASILY be the result of CHANCE? Of course they could be chance! Imagine you did the study again. Do you think you would get the same outcomes? No, sorry, **a small difference between small samples isn't statistically significant.** With only samples of five persons, you could easily get small differences by chance.

There are many kinds of tests of statistical significance. It depends on the kind of data you have---nominal, ordinal, interval, or ratio. **Basically, the test tells you the chances that results could be chance.** For example, a test might say, p = < .05 This means that the chances of getting the scores you got (e.g., the differences in the achievement scores for one group vs. another group) are less than 5 in a hundred. **The question is, CAN you live with that?** Is it okay to be wrong 5 out of 100 times? Would 95% confidence that the effect of a drug was real and not chance be high enough for you? How about the effect of a reading program? It would probably be satisfactory to have statistical significance at the .05 level. After all, you are only going to put the new program in once. The odds are 95 out of 100 in your favor. But if you used the program in 100 school districts, the results might be chance---not the result of the program---5 times.

Here are some resources on statistical tests.

http://www.graphpad.com/articles/interpret/principles/stat_principles.htm

Citation: H.J. Motulsky, *Analyzing Data with GraphPad Prism*, 1999, GraphPad Software Inc., San Diego CA, www.graphpad.com.

http://www.itl.nist.gov/div898/handbook/prc/section1/prc13.htm

http://www.surveysystem.com/signif.htm

http://en.wikipedia.org/wiki/Statistical_significance

http://www.statpac.com/surveys/statistical-significance.htm

Correlation

Let's say you are doing research that is looking to see IF there is a causal relationship between, say, how often teachers provide immediate and specific praise (input, independent variable, intervention), and the rate at which students give correct responses (outcome, dependent variable, effect). You have a pool of 50 children in fourth grade. The 50 children are assigned at random to two classes: Experimental group (teacher gives immediate and specific praise---"I love the way you answered with a full sentence!"---after almost every correct response); Control group (teacher gives delayed, general praise after one out of four correct responses. "Good job."). Here are the data.

```
Rate          Control group
30|
25|
20|                                      *              *
15|     *    *  *      *      *  * *    * *   *   * * *  *    * *
10|    *  *  *        *  *  *    *  *    * *         *       *
 5|   *        *
    ---------------------------------------------------------------
             5        10        15       20     25     30      35
                        DAYS
```

```
        Experimental group
                                                                   *
  Rate                              *           *  *    *  *
  30|                            *    *    *  *    *      *
  25|          *   *  *      *  *  *      *   *
  20|             *  *   *  *    *         *
```

```
15|        * *    *
10|
 5 |  * * *
    ----------------------------------------------------------------
             5         10         15         20      25      30        35

                        DAYS
```

Let's draw a **best fit line** through the data points.

```
Rate              Control group
30|
25|
20|                                                      *                         *
15|    *       *    *          *          *   * *      * *     *     * * *    *        * *
10|  * * *          * *   *    * *     * *          *             *
 5|  *          *
   ----------------------------------------------------------------
             5         10         15         20      25      30        35

                        DAYS
```

```
           Experimental group


  RATE                                                                               *
  35|                                              *                   *  *     *  *
  30|                                        * *      *   *    *          *
  25|            *   *  *        *  *  *      *    *
  20|               *  *   * *     *                  *
  15|        * *    *
  10|
   5 |  * * *
    ----------------------------------------------------------------
             5         10         15         20      25      30        35

                        DAYS
```

Notice that there is almost no change in the rate of correct responses in the control group. The rate begins at around 12 correct responses per lesson, and ends at around 15 correct responses 35 days later. However, **there IS an increase in the rate of correct responses in the Experimental group.** The group began with 6 correct responses per lesson and rose to about 35 correct responses per lesson 35 days later. **But how STRONG is the relationship between timely, specific praise and correct responses?** *How accurately (closely) does the number of days students receive timely and specific praise predict the number of correct responses **on that day**?* Well, look at the plotted data for the experimental group. If the correlation between praise and correct responses (if the prediction of correct responses from knowledge of days of praise) was perfect (100% accurate), then all of the data points would be right on the best fit line. But they aren't. This means that if on Day 20, you predicted 30 correct responses (as the line says), you would be off by 5 responses. The actual number of correct responses on Day 20 was 35. Check some of the other data points. What does the line predict for a day, and what is the actual number for that day?

So, does knowing the day enable you to predict rate of correct responses better than if you pulled a number out of a hat? Yes. Why? Because there IS an association (correlation = CO-relation) between days of praise and correct responses.

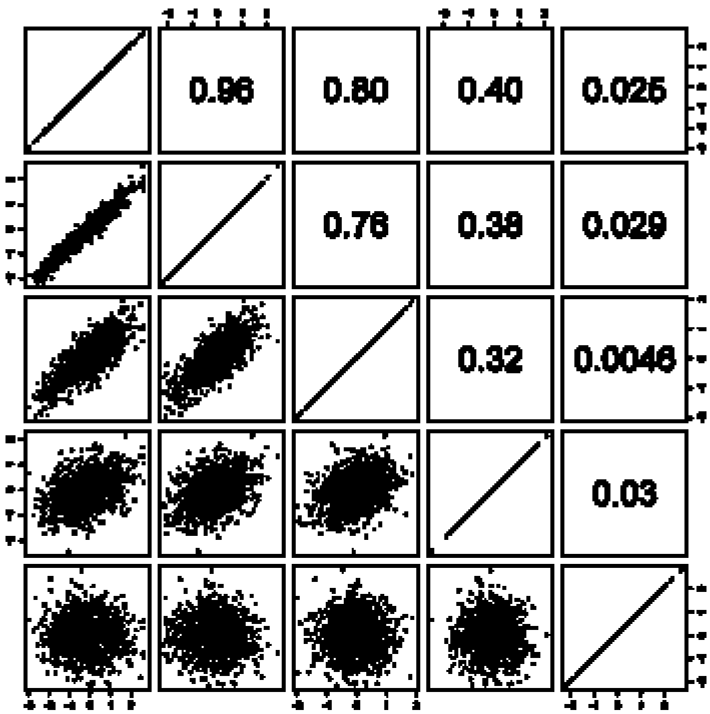
There are statistical techniques that tell you just how strong the relationship is. The number they give you is the "correlation coefficient." The table, below, is from Wikipedia. It shows the shape of a line, and it shows data points around the line. The numbers to the right are the correlation coefficient. For instance, the top left plot shows data points almost right on the line. This means that the correlation between one variable and the other is very high = .96.

In the second line down, the correlation coefficient is .76. Notice that there is more variation. The same spot on the across axis is associated with several different values on the up axis.

The correlation in the third row down is even weaker. Notice that any value along the across axis is associated with MANY values along the up axis. The coefficient is .32.

In the fourth row down, there is hardly any association at all. And the coefficient is .03.

The fifth line shows zero relationship. Knowing the value on the across axis does not give you **any** information about what the values on the up axis might be.

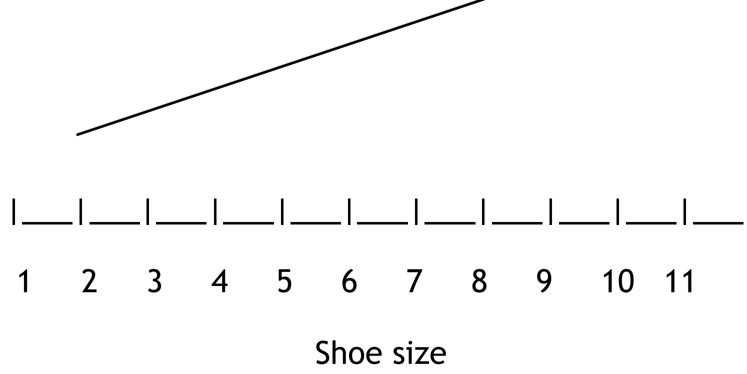


http://en.wikipedia.org/wiki/Correlation

Examine the graph, below.

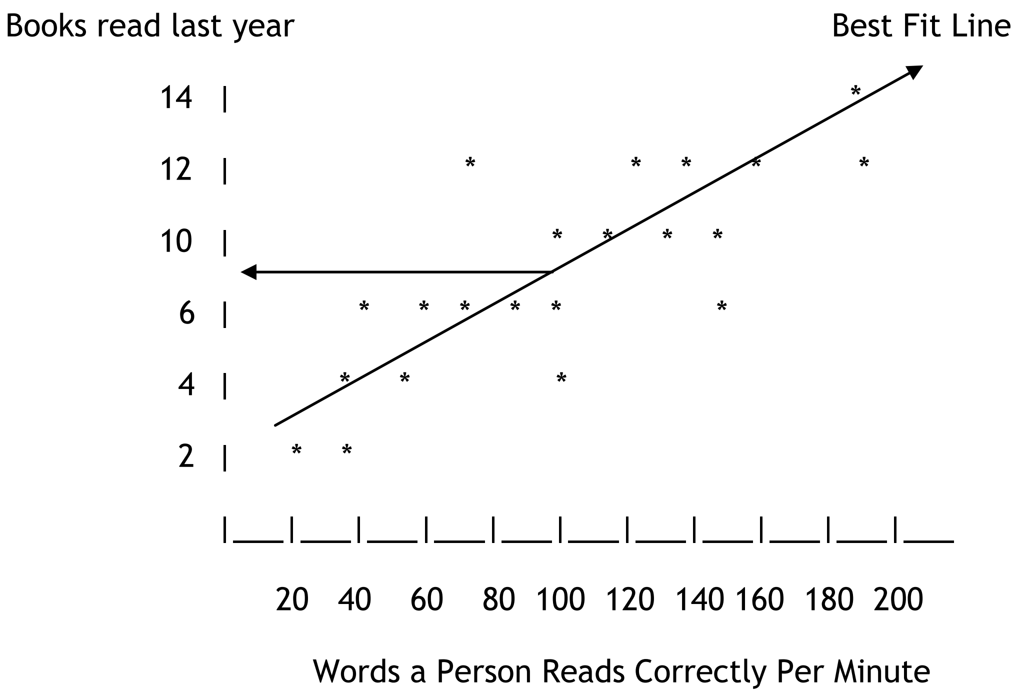
Books read last year

```
50 |
40 |                *                *     *
35 |                      *      *          *
30 |         *   *  *  *            *       *
25 |    *   *        *        *       *      *
20 |    *     *      *  *  *      *  *      *
```



Is there a trend here? Yes, people with tiny feet (infants) don't read much. And when people get older—and their feet get bigger---they read more. But some people with big feet hardly read any books. So, how well does shoe size predict the number of books a person reads? **How strong is the association?** Look at the table above from Wikipedia. The plot above looks like the fourth row down on the Wikipedia table. A correlation coefficient of .03. Almost nothing.

Here's another graph.



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. You can see that the higher the fluency, the more books persons read. Fluency IS correlated with, and it predicts, the number of books read.

Notice that the best fit line does NOT **connect** the plotted data points in a zig-zag line. It **cuts through them so that there are about as many above it as below it.**

Pick a value along the across (input, predictor) axis. Say, 100 words per minute. The best fit line predicts that persons reading at 100 words per minute will read how many books?.... (See arrow)…. 8 books. Now how many books did our teenagers reading at 100 words per minute **actually** read?... 4, 6, and 10. We predict 8, but we get a range from 4 to 10. This is PRETTY strong. Check the Wikipedia table. Which plot does our book plot look like?... I'd say the third row down. The correlation coefficient is .32.

Here are more resources on correlation.

http://www.neatideas.com/cc.htm

http://www.surveysystem.com/correlation.htm

http://ssed.gsfc.nasa.gov/lepedu/IA-CorrCoeff.html

http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

Review

1. Propositions
   a. are statements about which variables cause other variables.

b. are statements that connect variables in a theory.
c. are neither true nor false.
d. must not use negative terms.

2. Which of the following diagrams best shows the following theory? "Accurate comprehension of written text requires accurate decoding of text. But it also requires fluent or rapid decoding."

a. Comprehension → [If Fluent decoding] ← Accurate decoding
[Independent variable] [Dependent variable] [Intervening variable]

b. Fluent decoding + [If Accurate decoding] → Comprehension
[Independent variable] [Intervening variable] [Dependent variable]

c. Accurate decoding → [If fluent decoding] → Comprehension
[Independent variable] [Intervening variable] [Dependent variable]

d. Accurate decoding → [If Comprehension] → Fluent decoding
[Independent variable] [Intervening variable] [Dependent variable]

3. Which of the following isn't a direct measure of running speed?
a. Speed of leg muscle contraction.
b. Place that a person finished in a race.
c. How many other racers a person passed.
d. All of the above.

4. Which of the following isn't required of a good operational definition?
a. It is precise.

b. It gives examples.
c. It is brief.
d. It is consistent with the conceptual definition.

5. Which of the following IS required of a good operational definition?
   a. It is relevant to the setting.
   b. It excludes irrelevant examples.
   c. It uses concrete and clear terms.
   d. All of the above.

6. Here is a set of scores. What is the modal score?
   67% 50% 72% 89% 72% 85% 95% 90% 67% 72% 44%
   a. 72%.
   b. 85%.
   c. 67%.
   d. none of the above.

7. Here is a set of scores. What is the median score?
   44% 50% 67% 67% 72% 72% 72% 85% 89% 90% 95%
   a. Between 67% and 85%.
   b. 85%.
   c. 67%
   d. 72%

8. Here is a set of scores. What is the mean score?
   44% 50% 67% 67% 72% 72% 72% 85% 89% 90% 95%
   a. 65%
   b. 73%
   c. 85%
   d. 77%

9. Which of the following isn't an example of triangulation?
   a. A researcher uses both a standardized test of reading fluency and has each student read a passage from the course materials, while the researcher counts correct words read per minute.
   b. A researcher uses two different tests to measure the same variable.
   c. A researcher administers the same test several times.
   d. A researcher uses both an objective measure of teaching proficiency (e.g., observes and scores teachers on a form) and a pencil and paper test that asks questions that measure what a teacher knows of proficient teaching.

10. Which of the following is an example of an inter-observer reliability check?
    a. Different observers score different things.
    b. The same observer scores different things.
    c. Different observers score the same things.
    d. The same observer scores the same things.

Drawing Causal Inferences

Whether it is inductive theorizing and research (looking for empirical relationships) or deductive theorizing and research (testing hypotheses), we are likely to assert causal/functional propositions (i.e., to make causal inferences). However, as David Hume shows in *A treatise of human nature*, we cannot see causation directly. We only **infer** causal connection. Somehow, we feel more comfortable with the proposition "X causes Y" when the inference is drawn under certain evidentiary conditions, which are as follows: 1) evidence that the alleged cause preceded the alleged effect ("temporal priority"); 2) empirical evidence that the alleged cause and effect occur together ("contiguity"); 3) logical evidence that ties them together ("constant conjunction"); and 4) evidence that alternative explanations are implausible. Let's examine each of these criteria.

Alleged Cause Precedes Alleged Effect. Consider the following assertion. "An increase in teachers' authority to make curricular decisions (independent variable) fosters an increase in teachers' attachment to their school." This proposition (it could be an empirical generalization from research) seems plausible **only** if there is **evidence** that teachers' authority to make curricular decisions **preceded** an increase in teachers' attachment to their school. Evidence of temporal priority might be supplied by observation, experimental control, and/or commonsense reasoning (e.g., it isn't likely that a house burned down and then someone smoked in bed).

Empirical Evidence of Association. The inference that an increase in teachers' authority to make curricular decisions fosters an increase in teachers' attachment to their school, is more compelling if we have data showing that these two variables **changed in close succession** (V->Y: a proximal relationship) or in a sequence of variables that changed in close succession (V->W->X->Y: a distal relationship), and in the order asserted. Similarly, we can conclude that a family training program produced beneficial effects only if we have evidence of change in families and evidence that family members attended meetings, understood what was presented during meetings, and read and understood materials.

Evidence Provided by Inductive Logic. Logical evidence is obtained by designing research, analyzing data, and interpreting findings such that we can apply one or more of John Stuart Mill's methods of inductive inference, as described in his *A system of logic*. These methods include: concomitant variation, agreement, difference, joint agreement and difference, and residues.

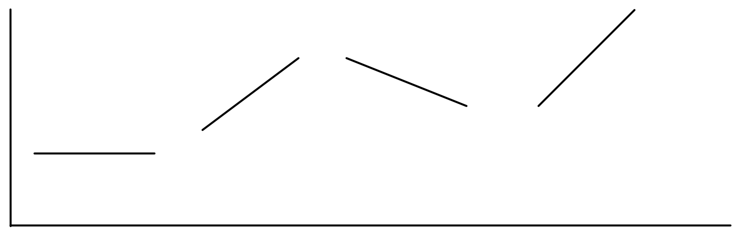
1. The method of concomitant variation. If two variables are changing with respect to one another (e.g., both are increasing, both are decreasing, or one is increasing and the other is decreasing) while everything else remains at about the same level, then we have logical evidence that one variable is a cause or an effect of the other (or they are both being changed by a third variable.)

For instance, an experiment was conducted to see how to increase the time on task in a student with severe attention difficulties. During the first experimental period (A1 or Baseline), the teacher was asked to go about her business and teach as usual. An observer ran a stop watch, recording when the student was on task. In the next experimental period (B1), the teacher was coached to reinforce the student periodically (with encouragement) when the student was on task.

In the third experimental period (A2), called a "reversal," the teacher was asked to do what she used to do during first A period. And during the final period (B2), she was asked to go back to reinforcing on task behavior.

Let's say that we graph the number of minutes on task per 30 minutes lessons. When the student received little reinforcement (the A periods), time on task was low; when the student received a lot of reinforcement, time on task went up. Since nothing else in the classroom was changing along with changes in the teacher's responses to time on task, it is plausible to infer that changes in the teacher's responses somehow caused **concomitant** changes in the student's time on task.

A1 B1 A2 B2



2. The method of agreement. Imagine that we study twenty failed school reform efforts. Each school and each reform effort was a **different configuration** of variables (e.g., school size, socioeconomic status of school, location, teacher-student ratio, speed of reform). Despite these differences, however, **all of the schools and failed reform efforts had one thing in common**--staff did not fully understand and were not fully committed to the mission or the reform plans. Since nothing else in the schools and plans was common across the schools, it's reasonable to infer that **the way in which they "agreed" (i.e., were the same) was the cause of the failed reform efforts.**

3. The method of difference. Mill's method of difference is the form of inductive logic used in the typical pre-test, post-test, experimental-group, control-group study. Let us say that we have a pool of 50 families whom we **randomly** assign to two comparison groups. One group receives written materials, ten weekly group meetings, and weekly home visits aimed to improve family interaction and home teaching. The second group receives written materials only. We compare pre-test and post-test scores on the quality of family interaction and home teaching. Families in the first group have significantly larger pre-post-test differences. What can we infer? Since we **randomly assigned** families to the two groups, any personal and family differences that might have accounted for improvement or lack of improvement (e.g., religion, support network, expectations of success, initial teaching skill) **had an equal chance of being in each group**. Therefore, ***we can assume that the groups were fairly similar on these extraneous factors***. (Of course, we could also measure those factors that we think are important and see how similar the two groups actually are.) **Since the only other *systematic difference* between the two groups (which we know about) was group meetings and home visits, it seems likely that these two features of the training made the difference in the amounts of improvement.**

4. The joint method of agreement and difference. This method combines the methods of agreement and difference. Let's take the above research on family training one step farther. We compared pre-post-test scores of families in the two groups which systematically differed only on whether they received written materials or received materials, meetings, and home visits. We used the method of difference to infer that the meetings and home visits accounted for the difference in improvement. Now imagine that, in addition, we obtain a large sample of families who **differ in many ways** (income, ethnicity, education, etc.). In each family we examine the quality of family interaction and teaching (dependent variables). We also examine whether each family reads materials on interaction and teaching (e.g., books, magazines), is part of some kind of group in which family interaction and teaching are discussed, or receives any in-home assistance or support (e.g., from relatives or other families) (independent variables). If we find that families who attend family-oriented meetings and receive home assistance also have higher quality family interaction and teaching, then we have logical evidence through the method of agreement that these variables make a difference. In summary, **the combined use of the methods of agreement and difference provides compelling evidence.**

5. The method of residues. Imagine a situation in which some phenomenon (Y) might be explained by four factors. We may be able plausibly to infer the one factor that is the cause through a **process of elimination**. If we know that factor 1 is a cause of Q, factor 2 is a cause of R, and factor 3 is a cause of S, then factor 4, the only one left, is likely to be the cause of Y. As Sherlock Holmes used to tell Dr. Watson, when you eliminate all of the other possible explanations, the one that remains, improbable though it may seem, must be the correct explanation.

Ruling Out Rival Hypotheses [See Extraneous Variables and Internal and External Validity, below.] Let's say we have satisfied the first three criteria for drawing a plausible and compelling causal inference. 1) We have evidence that the alleged causes preceded the alleged effects; 2) We have empirical evidence (data) that the two variables changed, and that they changed in the way that was asserted; and 3) We have used Mill's methods to provide logical evidence of a causal connection. Now we must show that **rival explanations are implausible**.

Consider the inference that children's rate of aggression changed as a function of change in teacher's responses to aggression vs. nonaggression. Surely it is possible that **other** variables caused some or all of the change in children's behavior. We must identify as many of these **extraneous** variables as we can and see if they provide plausible "rival" explanations. Below are some possibilities.

1. There were changes in some children's diets during the experiment (e.g., less sugar and less food additives). [Our data show that there were no such changes.]

2. There were changes in some children's participation in sports after school. Increased exercise calmed the children. [Our data show that only two children

increased their amount of exercise. This couldn't account for more than a small amount of change in the rate of aggression for the class.]

3. Maturation accounted for change in aggression and nonaggression. [It is unlikely that the children matured in the B1 period, regressed in the A2 period, and matured again in the B2 period--all coincidental with changes in the behavior of the teacher.]

4. During the A1 and A2 periods (when rates of aggression were high), the children were given harder tasks. Frustration was the cause of their aggression. [Our data show that the tasks were the same across all four periods.]

5. Some children were put on medication during the experiment. This caused a decrease in aggression. [Our data show that four children were put on medication during the experiment. However, two of these children were on medication during the A1 period (when the rate of aggression was high), and all four of the children were on medication during the A2 (reversal) phase, when aggression rose again. If we cannot say that medication decreased aggression during the A1 and A2 periods, it is unreasonable to think that medication worked during the B1 and B2 periods.]

6. The children's rates of behavior were really the same across the four experimental periods. The apparent changes were the result of measurement error or bias. [In fact, observers were trained to high levels of reliability before the experiment began. Their reliability was checked periodically during the experiment and was high. Moreover, observers were "blind" to the experimental periods and did not know what the hypotheses were.]

By showing that rival explanations are either false or implausible, it is likely that our explanation is correct.

Extraneous Variables and Internal and External Validity

Extraneous variables are variables that aren't a *planned* part of research; e.g., research on whether certain instructional methods are **associated** with higher achievement. Extraneous variables may **"interact with"** independent (input, intervention) variables (instructional methods) to produce an effect; e.g., teacher warmth may interact with how the teacher demonstrates a math routine, and yield greater attention and acquisition of skill. Or extraneous variables may produce an effect all **by themselves**; e.g., home instruction might increase math achievement. Therefore, change (or lack of change) in **dependent** (outcome) variables (e.g., math achievement) may be entirely or partly the result of extraneous variables, such as maturation, or other things happening inside and outside of school (e.g., siblings teach some students to read) or measurement error (students appear to read better because observers at the outcome assessment failed to count many errors) or bias in selection (e.g., if the experimental group has many bright students and the control group doesn't, that difference---and not instruction---may account for differences in achievement). **Findings and conclusions aren't valid, credible (believable), dependable, or good evidence for making decisions if researchers *can't rule out* the strong possibility that OTHER factors account for findings.**

Let's say you are using **official statistics** to see which kind of reading program produces the highest achievement in a school district. You use end-of-grade test scores to divide the district into (1) schools in which over 80% of students pass the tests; and (2) schools in which less than 80% of students pass the tests. Then you contact the schools and interview administrators and teachers to find out **how their schools teach reading**. But what if, in general, schools with the lowest reading achievement **also** have administrators who are so out of touch with school instruction that they really can't tell you how they teach reading? Your data on how their schools teach reading would **not** be accurate; your data would be **invalid**. In this case, the inaccuracy of the "observer" (the administrator) is an **extraneous variable** that muddies the picture. It's called "extraneous" because **it's not part of the process you are investigating.**

The Relationship Being Studied

Reading instruction → Measurement (and therefore findings) on Reading Achievement

^ ^ ^ ^ ^ ^ ^ ^ ^ ^
| | | | | | | | | |

Maturation	Measurement error
Home instruction	Biased sample
Teacher communicates high expectations	Experimental mortality (poor readers drop out)

Possible Extraneous Variables that Influence Findings on the Possible Relationship

Or, let's say you are conducting an **experiment** to see if peer tutoring will increase retention of math skills. Some students participate in peer tutoring (for a month) and others don't. You compare math skills before, right at the end, and every week for five weeks **after** the month of peer tutoring. Sure enough, students who participated in peer tutoring **do maintain** math skills more than students who did not receive peer tutoring. But what if more of the students who received peer tutoring also had **siblings who worked with them on math at home**, and what if this IN PART affected their retention of math skills? Help at home would also be an **extraneous variable** that muddies the picture, and makes your findings ("Peer tutoring seems to be effective.") invalid. In summary, many other things besides the hypothesized independent variables can account for findings. An important part of research is trying to **rule out the possibility that OTHER factors (extraneous variables) account for findings.**

"Internal validity" refers to how accurately the data and the conclusions drawn from the data (e.g., Change in X causes change in Y) represent what **really happened**. For example, looking at pre-test and post-test scores, it may seem that a training program increased teachers' skills. However, some of the difference between pre- and post-test scores may be the result of **measurement error**; during the post-test, observers wrongly scored some sloppy teaching as “proficient.”

"External validity" refers to how accurately the data and your conclusions drawn from the data (e.g., Change in X causes change in Y) represent what goes on in **the larger population**. For instance, if a sample of teacher-trainees is biased in some way (e.g., the sample contains a higher proportion of motivated trainees than is found in the general population of potential teacher-trainees), then findings from the sample may not apply to (won’t describe) the general population.

Note that findings and inferences may have internal validity but not external validity. That is, findings and conclusions may accurately represent what was found in the sample studied, but may not apply to **other** samples. However, if findings and conclusions don’t have **internal** validity, then they surely don’t have external validity either.

The factors that can weaken internal and external validity are called **"extraneous variables.”** Maturation of study participants is an example. Change in children's skills during instruction may reflect maturation of the nervous system and muscles as well as the effects of instruction. So, if the research hypothesis is that instruction will increase children's skills, the **"*rival* hypothesis"** is that maturation will increase children's skills. That’s why it is important to identify possible extraneous variables (sources of "contamination"). You can then design research to weaken or eliminate the effects of these variables, or you can analyze the data to determine what effect the extraneous variables have had. For example, if you use an experimental and control group, and if you created the two groups using the method of **randon allocation,** then you weaken the rival hypothesis of maturation---since children in both groups have an **equal chance** of improving as a result of maturation.

Extraneous Variables That Are Threats to *Internal* Validity

1. Instruments don't measure what they purport to measure. In other words, the findings aren't valid. For instance,

a. The dependent (outcome) variable is reading proficiency. However, that isn't what the researcher is measuring. Instead, the researcher is measuring behavior such as turning pages, naming parts of a book, holding books properly, memorizing words, and guessing what words say. If the researcher is "testing" a new method of reading instruction (a method that does NOT work), this method will APPEAR to be effective **because the researcher isn't measuring reading at all**. Consumers should expect researchers to use **standardized validated** methods and instruments, or expect researchers to carefully define variables, and then develop valid measures based on these definitions.

b. **The measurement method or instrument has *not* been tested for reliability**; that is, different observers or testers observing the same thing would NOT get the same scores. If a method or instrument isn't of known high reliability, then a group receiving an intervention may **<u>appear</u>** to have made a lot of progress between pre-test and post-test, but only because the post-test scores were ***wrong*.**

   Therefore, consumers should expect researchers to use methods and instruments with known high reliability, and should expect researchers to check that observers and testers produce reliable data before a study begins, and periodically during a study if repeated measurement is used.

c. **Data that *should* be OBJECTIVE** (e.g., counting how often teachers properly correct student errors) **are in fact <u>subjective</u>---opinions, impressions.** These subjective data ("I learned a lot!" "Training was excellent." "I am confident that I can properly teach the five reading skills.") can't be used to determine if a program or method is effective or if teachers are proficient. Why? For the same reason that you can use subjective opinions to determine if a medication is effective. If a drug works, there will be objective changes in the **body**. If a program works, there will be objective changes in student **behavior**. Opinions don't measure proficiency; they measure feelings. Also, opinions and feelings and impressions change—and therefore aren't reliable indicators of hard facts of proficiency.

2. History. History includes events **in addition** to the independent variables under study, that occur **between** one measurement and another (e.g., between a pre-test and post-test). For example, in testing the effects of an exercise program on psychological well-being following heart attack, some participants joined a church, or received additional social support, or changed jobs. These extraneous (history) variables may account for some of the differences between pre- and post-test scores.

To weaken history as a rival hypothesis, researchers should use **equivalent experimental and control groups** (created by random allocation or matching). Since the groups are, logically, likely to have the **same** historical variables happening between pre- and post-test, differences in the outcomes aren't likely to be the result of history.

3. Maturation. Maturation refers to changes that ordinarily **occur with time** (e.g., strength, increasing knowledge). For instance, let's say a new method to increase children's attention span is tested in an experimental intervention. And let's say that most children are more attentive two months later, during the post-test. The experimenter may think this improvement is the result of the method, but the rival hypothesis is simply that the children became more mature, and THAT increased their attention. To handle the extraneous variable of maturation, researchers use equivalent comparison groups, or us use experimental designs in which the experimental group serves as its own control (e.g., the equivalent time samples design).

4. Testing. This refers to the effects of taking one test on the results of a later test. For instance, improvement in scores might reflect decreasing fear of being tested, or figuring out what kinds of answers are correct.

Testing can be controlled in part by using different versions of the same tests and by using comparison groups in which one group does not receive a pre-test.

5. Statistical regression. A person's performance of any task varies within a certain range. On the average, you may be able to do 10 pull-ups, but on a particular day you may do 8, 9, 11, or 12. In fact, there may be days when your performance is quite **unusual**--you can barely do 5 pull-ups, or somehow you manage to do 18. However, if you did pull-ups the next day, and the day after that, your performance would probably **regress (move) to the mean**, or your average performance.

In research, a group's pre-test performance might (by chance) be **unusually** high or low; some people had a good day or a bad day. On later testing, the group's performance regresses to the mean (i.e., is more usual). The researcher may mistakenly treat differences between pre- and post-test scores as the result of an intervention ("They improved.") or as the failure of an intervention ("They got worse!"), when in fact, the group merely turned in its average or usual performance.

The rival hypothesis of statistical regression can be partly controlled by using equivalent comparison groups, since the possibility of unusual scores applies equally to the groups.

6. Selection bias. In research using comparison groups, some participants in one group may be different from those in the other group(s) in ways that affect performance. For instance, an experimental group may do much better on a post-test than the control group, ***not*** because the experimental intervention was effective but because more of the experimental group members figured out how to take the test (See number 3 above.). Similarly, the pre-test/post-test differences between the experimental and control group may be small, suggesting that the intervention **did not work**. However, in fact, the control group contained many people who **WERE** likely to change as a result of maturation or some historical factor, and so they gotr high scores even though they received no intervention.

This source of invalidity can be handled, in part, by random allocation of participants to comparison groups. This way, all possibly biasing factors have an equal chance of being in both groups.

7. Experimental mortality. This refers to the differential loss of participants from comparison groups. For example, an experimental intervention may appear to work only because participants with whom it **was not going to work *dropped out***. Similarly, an intervention may appear to work no better than nothing at all, only because people in the control group who would have gotten **WORSE** over time dropped out, leaving people in the control group who improved. Thus, the control group scores about the same as the experimental group.

The rival hypothesis of experimental mortality can be partly controlled by using equivalent comparison groups, since the chances of dropping out should be about equal in the two groups.

9. Causal time order. Here, participants began to **change *prior* to an intervention, but the researcher does not know this**. It only appears that the intervention is the cause of the change.

```
|         |                                  *
|         |                             *
|         |                   *    *
|         |         *    *
|         |    *
|    *    |
|         |
|         |
|____|____|____|____|____|____|____|____|____|_____
```

Pre-   Intervention
Test

This is what is really happening.

```
|            |                          *
|            |                      *
|            |              *   *
|            |      *   *
|            |  *
|         *  |
|      *     |
|   *        |
|__|__|_|_|__|__|__|__|__|__|__|__|__|__
```

Repeated   Intervention
Pre-tests

Notice that change during the intervention is just a *continuation* of what had already begun.

A partial solution is an extended series of repeated baseline or pre-intervention observations, to assess the stability of performance before an intervention. If the "baseline" or pre-test scores are stable (mostly a straight line), and scores ONLY rise AFTER the intervention begins, you have evidence that the intervention is having an effect. For example

```
|            |                          *
|            |                      *
|            |              *   *
```

```
|          |      *   *
|          |   *
|          |
|          |
|   *   * * |
|__|__|_|_|__|__|__|__|__|__|__|__|__

Repeated               Intervention
Pre-tests
[Stable]
```

10. Diffusion or imitation. Here, part of an intervention given to an experimental group is used by members of the **control group**. Thus, the intervention does not appear to make much of a difference, because both groups have changed. For example, families in a training program lend materials to friends in the control group.

One way to try to control this is to make sure that members of the comparison groups don't know one another. Another method isn't to tell participants what group they are in---a **single blind study**. However, this may pose ethical problems. Still another method is to use delayed-intervention control groups (so that members of the control group may be more willing to wait).

11. Compensatory rivalry. Knowing they are in a control group, some participants try to change on their own. Improvement in the control group may be mistaken to mean that the intervention is no better than no intervention. One way to handle this isn't to tell participants which group they are in. This is called a "single blind" study.

12. Demoralization. Knowing they are in a control group, and not receiving an intervention that they want, some members of the control group look worse over time than they otherwise would. This may result in differences between the experimental and control group being mistaken for the effects of the intervention. (Imagine the effects on their life expectancy if people with aides knew that they were in the control group of a drug experiment.)

A partial solution is to use a delayed-treatment design (rather than no-treatment design). Also, one could use alternative treatment groups rather than a control group.

Extraneous Variables That Are Threats to *External* Validity

Keep in mind that all of the threats to internal validity are also threats to external validity. Additional threats to external validity include the following.

13. Reactive or interactive effects of testing "Reactive" effects of testing means that a pre-test alone influences post-test performance. "Interactive" effects of testing means that a pre-test influences how people are affected by an **intervention**. If the performance of an experimental group after an intervention has been influenced by the pre-test, the findings (e.g., amount of beneficial change resulting from treatment) may not apply to the general population which isn't **likely to receive a pre-test**.

Therefore, it may be important to assess the effects of pre-testing itself. An experimental design call the Solomon Four-group Design is an effort to control this source of invalidity.

14. Interaction of selection bias and X (intervention) Here, a bias in the selection of the experimental group results in enough members of the experimental group being especially likely to be affected (or not affected) by X, so that the experimental group's post-test scores are higher than scores of the control group. But since samples in the general population aren't likely to have this bias, the results of the intervention with other samples may be less than in the experiment.

One way to handle selection bias is to use **random sampling** so that study samples are equivalent to the general population.

15. Interactive effects of experimental arrangements. If the performance of people in an experimental group was affected (positively or negatively) by certain features of the experiment, or by the fact that it was seen by them as an experiment, findings from the experimental group may not apply to samples from the general population who will receive the intervention in a **nonexperimental** setting. For instance, teachers in an experimental training program (which gives them a sense of being special) may change more than later trainees who simply receive a course on the same material. There is no way getting around this one. The more you control a situation so that you get valid data, the LESS the situation is like real life, and therefore, the results you got in the contrived setting may not happen outside of it. However, you CAN TEST THAT very hypothesis by replicating the research in more and more natural settings, and see if the results remain about the same.

References

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.

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.

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.

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.

Walberg, H.J. (1990). Productive teaching and instruction: Assessing the knowledge base. *Phi Delta Kappan,* February, 470-478.

Vocabulary

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 trhe 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 a study that differ in some way (e.g., curriculum, whose effects or outcomes are being investigated). In an experiment, one of the comparison groups is always 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. The comparison groups in studies examining cause and effect relationships should be as similar as possible, so that the only significant difference is the intervention (e.g., curriculum) that each received.

Control groups. A control group is a group in an experimental study 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 aren't 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.

Conceptual definitions. Conceptual definitions are broad. They are like a search light that shines on a general area.

Operational definitions. Conceptual definitions aren't precise enough to use to determine what to measure, and instruments for measuring. Operational definitions give examples (based on the conceptual definition) of EXACTLY what you would see or hear.

Disaggregation of data. Examining a data set by separating (disaggregating) data on the basis of some factor, like student gender or ethnicity (this is the opposite process to aggregation of data where data are put together). 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. [See Aggregate data.]

Diverse Learners. Learners in some group that possess varying levels of some important characteristic. The term is often used to refer to wide variation in economic status, background knowledge, ethnicity, race, or class that appear in many classrooms and school. However, the term is equally applicable to differences in any relevant factors (e.g., ability, interest, aptitude, etc.).

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.

Randomization, or random allocation. This is the preferred 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.

Matching. Matching is a second 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. However, a researcher can't know ahead of time all of the variables that describe a group could affect the outcomes. Therefore, the researcher can't be sure that the groups really are equivalent. For example, the groups may differ in help students receive at home, and THIS (not the intervention) may be what makes one group achieve more than the other. This is why randomization is preferred.

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 aren't 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 don't 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 isn't 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 don't 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 isn't 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 don't have a stake in the outcomes of the research. For example, it isn't 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 don't receive free and reduced lunch. This nominal (name) measurement indicates difference in family income, but it isn't 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 aren't 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 aren't 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 aren't 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 isn't 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 aren't 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.

Weight

50 |

40 |

35 |

30 | *

25 |

20 |

|___|___|___|___|___|___|___|___|___|___|___

1 2 3 4 5 6 7 8 9 10

→ Age

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 isn't 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 shouldn't 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.

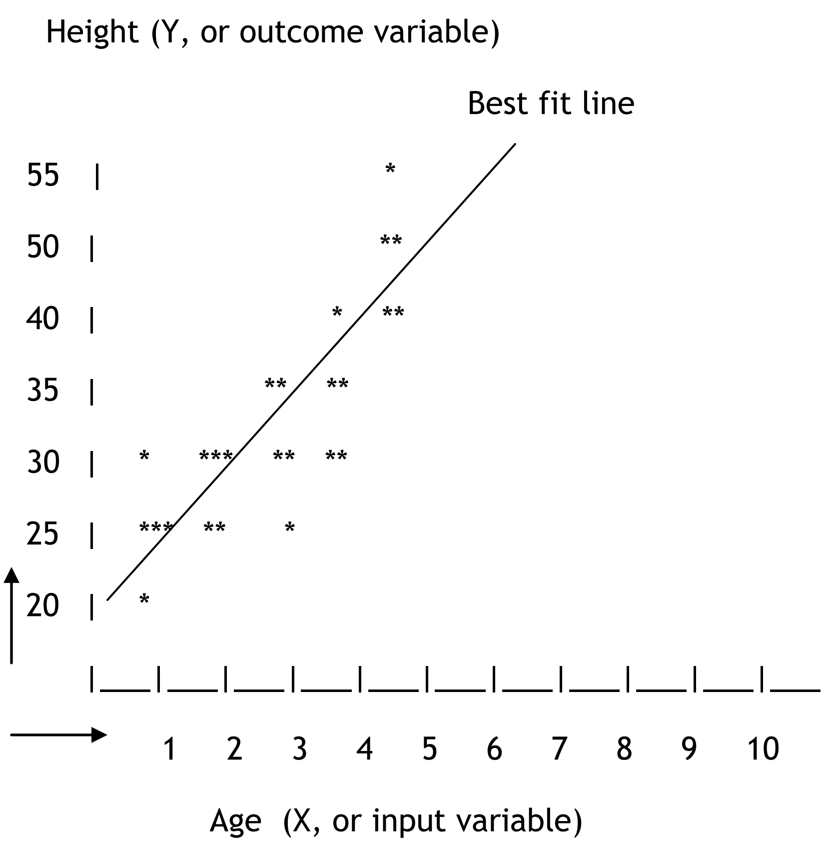
Range. Range means the spread of scores from lowest to highest. For example, a group of persons ran as far as they could. The shortest distance run was 1 mile. The longest distance was 40 miles. So the range is from 1 to 40 miles. It doesn't matter if only one person ran 40 miles or if five persons did.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

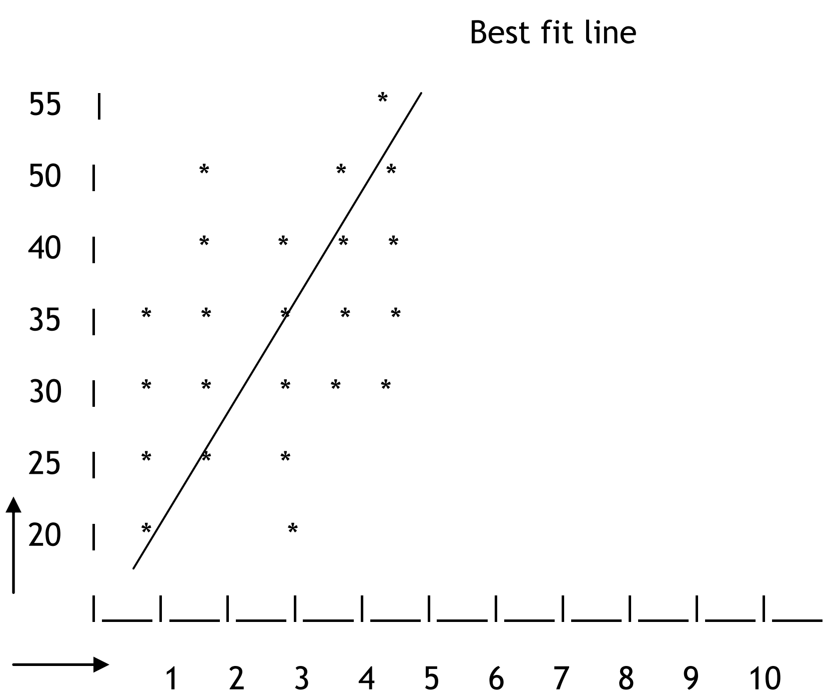
Representative Sample. A sample whose characteristics (e.g., percentage of persons of different ethnicities, social classes, sexes, skills) are similar to the characteristics of the population to which the findings are relevant. If a research sample isn't representative of the relevant population, you 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 isn't 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 don't 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 isn't 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

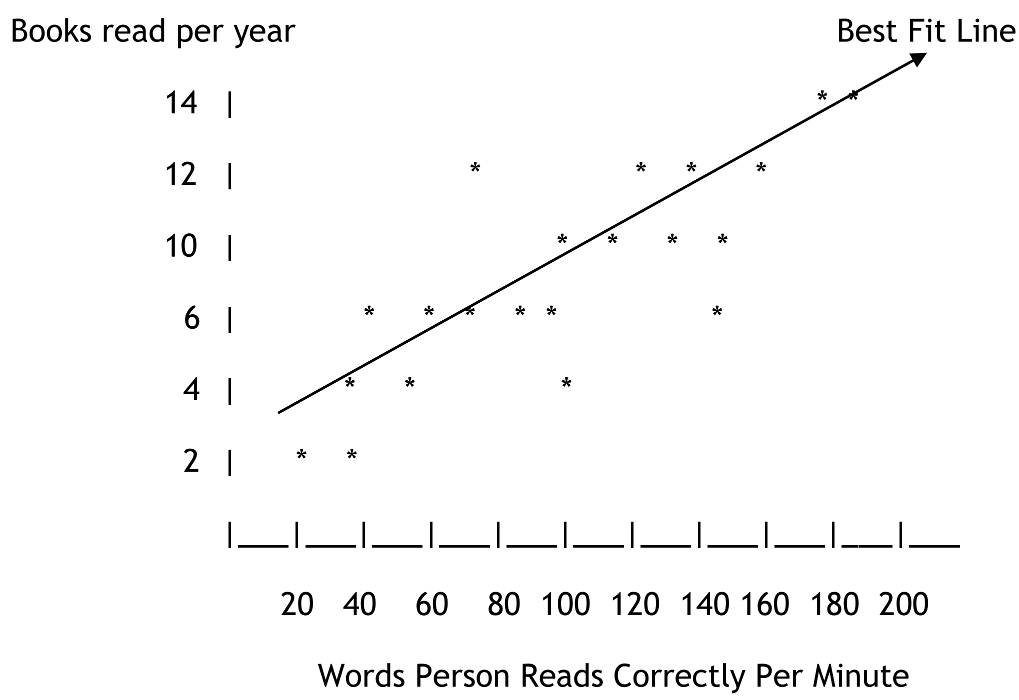
```
14 |                                             *  *
12 |                  *            *   *    *
10 |                         *  *    *   *
 6 |          *   *  *   *  *            *
 4 |        *    *           *
 2 |    *   *
   |____|____|____|____|____|____|____|____|____|____|____
         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 isn't a measure of (isn't 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 isn't 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 aren't 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 curriculum materials	→	[Given the degree of teacher proficiency using the materials]	→	Student achievement

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.

All Assignments

Assignment 1. Examine an historical document.

http://www.yale.edu/lawweb/avalon/avalon.htm

a textbook, or instruction on the internet. Identify and write at least two examples of

1. Simple facts.
2. Lists.

3. Higher order concepts.
4. Rules or propositions.
5. Routines. Note that a theory or logical argument would quality as a routine.

Assignment 2.

Get a textbook in some subject, or a book written on a subject---a knowledge system. Select a paragraph or two**. Identify important things to teach and list them. State what kind of knowledge is each "thing"?** There may be facts and lists; higher-order concepts that the book defines or that you will have to define for your students; rules stating how one set of things goes with (e.g., causes) another set of things; and maybe routines, such as a theory (a sequence of statements that lead to an explanation). **Do this with a second subject, or knowledge system.** You'll find the same thing. *Everything you want students to learn will boil down to the five kinds of knowledge*.

Assignment 3.

Here are data on two **variables**---*classes of stuff whose examples have varying values*. Each row (an example of a general rule/connection) says "When you have this much _______________, you also have that much ________________."

Use inductive reasoning to figure out (induce) a general rule about "this much" in relation to "that much."

State the steps that you used. Did you compare and contrast examples?

	Average Temp	No. of Suicides Monthly/1000
Jan	36	68
Feb	39	80
March	43	86
April	50	102
May	57	105
June	63	100
July	66	100

Aug	65	82
Sept	60	74
Oct	52	70
Nov	43	66
Dec	38	61

Rule = (When, the more, the higher) ________________, (then, the more, the less, the higher) __________________.

Now make a deduction from the rule. If ________________ goes from _______ to _________, what do you predict will happen to ___________? State the deductive reasoning as a syllogism, using the lines below.

[If it is true that] When _________________, then ____________________.
And if ___________________________.
Then, ___________________________.

Assignment 4.

Here are some examples. Try to fill in the blanks. What do YOU think important intervening variables might be?

Teacher creates cooperative learning groups →	[If teacher…..] →	Accomplishment of group tasks
[Independent variables]	[Intervening variables]	[Dependent variable]

Teacher establishes → classroom rules	[If …………………..] →	Students cooperate with rules
[Independent variables]	[Intervening variables]	[Dependent variable]

Assignment 5.

Here are examples of conceptual definitions**.** *Think of operational definitions for each one.* Remember, the operational definition has to say the say the same thing as the conceptual definition, but it is more precise; it gives examples. Also, operational definitions **depend on the situation**. For example, part of an operational definition of aggression might be hitting, but NOT if you are talking about the sport of boxing!

Conceptual definition	**Operational definition**
Aggression is behavior that is Intended to cause injury	[Aggression on an elementary school playground]
Reading fluency Fluency is a feature of performance: The combination of accuracy and speed.	Second grade reading fluency http://reading.uoregon.edu/flu/flu_benchmarks.php
Proficient instruction is instruction that is carefully designed and delivered and	You want a hint? NO! Let's see how tough you are.

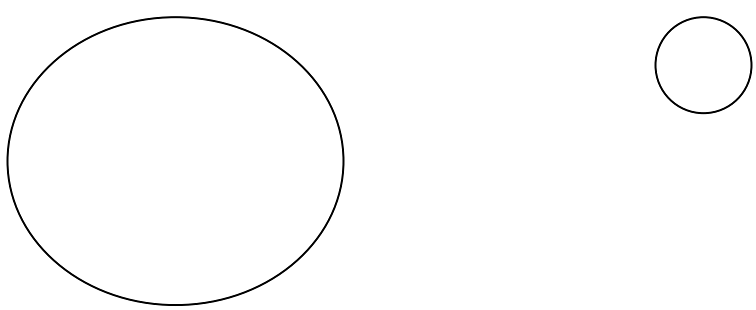
contains features that gain and
sustain student attention and
participation.

Assignment 6.

Write and diagram categorical propositions regarding the following sets of two variables: (1) things fostered by all skilled teachers and achievement in students; (2) successful school reform efforts and social systems in which members **don't** have a shared mission; (3) adults with antisocial personalities and children who receiv**ed** harsh discipline (all or some?); (4) effective leaders and persons who are trusted.

Here are some hints. Take number 1. Which is the larger category---achievement in students or things fostered by all skilled teachers? Which category has more stuff in it? Do skilled teachers foster achievement? Yes. Is that the ONLY thing they foster? No. Do they foster other things, too? Yes. So,

Things fostered by all skilled teachers. Student achievement.



Where does achievement go? Outside, inside, or partially inside things fostered by all skilled teachers?

The proposition would be: “All ______________________ foster ________________.”

Assignment 7. Concepts, Definitions, and Propositions

Following are excerpts that contain definitions and propositions. Find these and then state them in proper propositional and definitional form. Note that many propositions and definitions are implicit; e.g., the logical flow from proposition 1 to proposition 2 requires another (unstated) relationship or a definition. Also identify if there is reason to believe that the relationships are direct vs. inverse; uni-lateral vs. bi-lateral; involve necessary, sufficient, or intervening variables.
Yes, doing this WILL hurt your head. BUT you will be so much sharper!!!

1. ...a state is a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory. [Max Weber. "Politics as a vocation." 1918*]* *Is he asserting causation or is he defining a concept?*

2. No living being can be happy or even exist unless his needs are sufficiently proportioned to his means. [Emile Durkheim, *Suicide*. 1897]

3. If the state is to exist, the dominated must obey the authority claimed by the powers that be. [Max Weber. "Politics as a vocation." 1918] *Is he asserting that something is a necessary condition for another thing?*

4. ...the term suicide is applied to all cases of death resulting directly or indirectly from a positive or negative act of the victim himself, which he knows will

produce this result. An attempt is an act thus defined but falling short of actual death. [Emile Durkheim, *Suicide*. 1897] *Causation or definition?*

5. *If therefore industrial or financial crises increase suicide*, this is not because they cause poverty, since crises of prosperity have the same result; *it is because they are crises*, that is, *disturbances of the collective order*. [Emile Durkheim, *Suicide*. 1897] *Both causal and categorical propositions here*.

6. Where the State is the only environment in which men can live communal lives, they inevitably lose contact, become detached, and thus society disintegrates. [Emile Durkheim. *The Division of Labor in Society*. 1893] *Do you see a causal chain?*

7. There is the authority of the extraordinary and personal gift of grace (charisma), the absolutely personal devotion and personal confidence in revelation, heroism, or other qualities of individual leadership. This is charismatic domination... [Max Weber. "Politics as a vocation." 1918]

8. Quoting Esquirol (with whom Durkheim disagrees):
"Suicide shows all the characteristics of mental alienation." (p. 58)
"Mental alienation" means insanity. Which is the more inclusive category?

9. Average of Suicides per Million Inhabitants [Do you see a range of orthodoxy here?]

Average of Suicides per Million Inhabitants	[Do you see a range of orthodoxy here?]
190	Protestant States
96	Mixed States (Protestant and Catholic)
58	Catholic States
40	Greek Catholic States

So, you could say "The more….., the higher the…."

10.

Provs with Cath Minor (<50%)	Suicides/ Million Inhab	Provs with Cath Major (50-90%)	Suicides/ Million Inhab	Provs with More Than 90% Cath	Suicides/ Million Inhab
Rhenish	167	Low. Franc	157	Upp. Palatin.	64
C. Fracon.	207	Swabia	118	Upp. Bavaria	114
Upp. Franc	204			Low. Bavaria	19
	Ave. 192		Ave. 135		Ave. 75

What *empirical generalizations* can we draw from the above table?

"(S)uicides are found to be in ___________ proportion to _____

and in _______proportion to_________________________" (p. 153)

11. "(W)hen religious intolerance is very pronounced, it often produces an opposite effect. Instead of exciting the dissenters to respect opinion more, it accustoms them to disregard it." (p. 156)

12. "(A) religious society cannot exist without a collective credo." (p. 159)

13. "(T)he more extensive the credo the more unified and strong is the society." (p. 159)

14. "(T)he greater concessions a confessional group makes to individual judgment, the less it dominates lives, the less its cohesion and vitality." (p. 159) Causal sequence?

15. "Man seeks to learn and man kills himself because of the loss of cohesion in his religious society; he does not kill himself because of his learning." (p. 169)

16. "(T)he desire for knowledge wakens because religion becomes disorganized." (p. 169)

17. "(T)he density of a group *[rate of interaction]* cannot sink without its vitality diminishing." (p. 201)

18. "Excessive individualism…frees man's inclination to do away with himself from a protective obstacle… (p. 210)

19. "...they are crises, that is, disturbances of the collective order." (p. 246)

20. "...more depressed and anxious pregnant teenagers, who perceive their social relationships to be less satisfying, and who have less knowledge of child development, have more negative expectations for their infants." J.M. Contreras et al. (1995). *Journal of Applied Developmental Psychology, 16*, 283-295. *Do you see the intervening variables?*

21. High mother support was associated with more secure infant attachment only for those adolescents living with partners." S.J. Spieker (1994). *Developmental Psychology, 30*, 1, 102-111. *Do you see an intervening variable?*