Chapter 2

Aspects of Validity in Quantitative Research

Gregg Schraw, *University of Nebraska-Lincoln*

I want to do three things within the next half-hour. First of all I would like to just give you a real quick overview of what John Creswell and I plan to do. We do have a master plan for this. Then I want to say a little bit about what cognitive psychologists see as the big hot spots in science research. I'm guessing there's quite a bit of overlap between your perceptions of science research and our perceptions of science research. Then I'd like to take a little bit of time to talk about some of the foundational assumptions of quantitative research and just give you some of my opinions. I'm assuming that most of you are pretty familiar with what I'm going to talk about.

Let me move to the first agenda item—that's John and I coordinating. John and I are both in the Educational Psychology program here at the University of Nebraska. I was originally trained as a statistician and worked as a statistician. I became interested in cognitive psychology, so now my line is actually kind of interesting. Half of my time is spent in cognitive psychology and I do a lot of research in beliefs and how they affect learning, as well as metacognition. I don't really do much research in the area of science education. The other half of my time is spent teaching advanced quantitative research methods. I'll talk a little bit about that. John's expertise is in qualitative research, so John and I collaborate quite a bit. We take one another's courses and we're thinking about doing some other projects together, along the lines of exploring mixed methodology and the foundations of qualitative and quantitative research. What we'd really like to do today is give you a smorgasbord of what goes on in terms of research options, both quantitative and qualitative, and I think John will speak a little bit, also, about transitions in education. I'm sure most of you are familiar with this but really up until about 15 years ago quantitative research was the main route for researchers, and its changed dramatically. For example, if you were to go to a conference like the American Educational Research Conference, which is held annually, you would see that probably the bulk of research, including science education research, is now qualitative in nature.

Hot Topics in Educational Research

I'd like to say a little bit about what some hot research topics are. I will focus on just three areas. Now again, I'm guessing that there's quite a bit of overlap. I mentioned that half my time is spent teaching cognitive psychology courses. For example, in the fall I teach advanced cognitive psychology. During the summer sessions I teach a masters level course which is kind of a panorama of cognitive psychology and we talk a little in there about science education.

Of the three big areas that I see in science education, the one that is far and away the biggest, really for at least at least 10 years, is the area of *conceptual change*. There is a lot of debate about whether conceptual change should be hot or cold. Let me just introduce these ideas and then I'll go back and say a little bit more.

Another big area is *expert-novice differences* in science education. In particular, I think the really cool research is being done in expert-novice differences in scientific reasoning. A big name in this area is Deanna Kuhn, if anybody's heard of her. She's at Columbia University. She does a lot of very interesting work and had a big seminal paper come out in 1989 (Kuhn, 1989).

The third area is *self-regulation*, and you are probably a little bit less likely to have heard about this. This idea comes from social cognitive learning theory—the work of Albert Bandura. Its been updated by some of Bandura's former students. Basically self-regulated learning theorists like to make a distinction between *skills* and *wills*. *Will* being the motivation, *skill* being the cognitive skills you actually need. There's a lot of talk about what kinds of skills are necessary to become self-regulated in science learning, or any other kind of learning, and how we can best present those skills in an educational context.

Conceptual change

Having said that, let me back up a little bit now to the first point, conceptual change. Probably the biggest name person in this area, and I don't even know if she's doing much anymore, is Susan Carey. She really kick-started conceptual change research in education in the mid-1980s. She wrote a very important book (Carey, 1985). The topic has been very hot since then. Conceptual change, at least as cognitive psychologists see it, deals with the issue of what happens to students when they enter your classroom with a whole bunch of knowledge that's inappropriate, or inappropriate beliefs, or (the term we like to use a lot) naive theories. You may have encountered those terms. Naive theories are just basically whacked out theories about, for example, how the universe works. I mean, anybody who's ever taught astronomy, or even knows a little bit about astronomy, (I'm assuming most of you know quite a bit—certainly more than I do) realizes that many students, including older students—middle schoolers or high schoolers—might have entirely inappropriate ideas about the physical nature of the universe. So the big question is: when that occurs what do we do to address that problem? How do we bring about conceptual change? How do we basically get into somebody's head and change their knowledge structure, change their knowledge organization? It is a tough task. There's a lot of debate. Just to cut to the quick, the debate seems to center around what is known as hot or cold conceptual change.

Hot conceptual change is the idea that you sort of hit them below the belt. You get them in there and you go out of your way to create a lot of cognitive disequilibrium. You make sure they understand how inappropriate their own understanding is. You sort of tear them down and then build them back up again. The idea of cold conceptual change is that you try and do it more systematically, in a more rational way, without trying to create too much cognitive disequilibrium. Generally, at least among cognitive people, the consensus seems to be that hot conceptual change is the way to go. This means that teaching has to have some shock component. One of your purposes is to make students acutely aware of how inappropriate their knowledge is, and that may be unsettling for students. That's the theory of conceptual change. There are a many other aspects involved, but unfortunately I just don't have time to go into it.

Expert-novice differences

A lot of work is done in the area of expert-novice differences. The bottom line, and it is pretty disappointing in my opinion, is that unless they have had formal scientific training through some apprenticeship approach in physics or biology or anywhere else in the sciences, or, for example have had statistical training, or training in methodology, either quantitative or qualitative methods, virtually everybody, kids and lay adults alike, is extremely poor at scientific reasoning. In fact, they're basically inept. It is pretty frightening, so that raises a lot of questions about, for example, what should constitute a reasonable education in the United States? For example, should students during high school ages be given formal training in scientific reasoning? I think there is a big outcry, at least among cognitive psychologists, for those kinds of skills and programs to be brought into a high school setting. The same holds true at the university level. I think a lot of people, particularly science educators in Curriculum and Instruction departments and Ed Psych programs, feel that science training has been pretty inadequate. There is a big movement to substantially increase the number of classroom hours, as well as the amount of applied experience, science educators get before they go out and teach. If they don't get that, they don't have a very good understanding of the scientific method and they are very poor at identifying what constitutes legitimate evidence or using that evidence to reason hypothetically. In other words, they can't test theories and things of that nature.

Again, if I had to refer you to one person it would be Deanna Kuhn. If you were to look in PsychLit or Silver Platter or those sorts of electronic databases you would see that she has many many articles. Probably her best is one she published in 1989 in *Psychological Review* , called "Children and adults as intuitive scientists" (Kuhn, 1989). *Psychological Review* is the best review journal in the area of social sciences. The article is very comprehensive. I realize the article is almost ten years old, but it is probably still the best, and it deals with these issues in detail. That is a big, big area of research.

In general, expert-novice differences have been very big in cognitive psychology for thirty years and cognitive psychologists are notorious for getting physicists to do think-aloud studies. They are interested in looking at how experts acquire a knowledge base, organize the knowledge base and use that knowledge base in a very strategic way. I guess I'll just leave it at that.

Self-regulation

The third area, self-regulation, is much more recent, but it is a very hot topic among cognitive people. There are a couple reasons reason it's a hot topic. First, the *skill* part. One is that in the last decade cognitive psychologists have decided that the best thing you can do in an educational setting, the best way to get a bang for your buck, is focus on strategies and improving motivation rather than trying to substantially change the knowledge base. The reason cognitive psychologists have sort of moved away from the knowledge base is just the realization that to really get an expert knowledge base in place is going to take probably five years. Maybe a little bit less, but it's going to take a long period of time. It's just too labor intensive, so cognitive folks have really shifted their emphasis to promoting intelligent strategy use. The reason is that you can get in there, educationally, and within a three or six month period you can teach a wide variety of very flexible strategies that substantially improve people’s information processing. The general idea with strategy instruction is that you don’t change what’s in the system very much, but you enable people to use what they already have much more efficiently.

Similarly, the motivation stuff, the *will* aspect of self-regulation. With motivation, what cognitive psychologists have found is (especially on the coattails of 10 or 15 years of strategy research) that: you go into classrooms; you teach all these great strategies; you give students a good knowledge base and you find out they don’t have the motivation or the will to use it. It is pretty disappointing. Cognitive psychologists have changed their attitudes very substantially over the last decade about the role of motivation. It is now seen as almost the place to start, and one of the most important components. As I just mentioned, in a self-regulated learning program, typically what you see is some emphasis on trying to impart a knowledge base as quickly as possible and then teaching a lot of specific strategies about how to do the kinds of tasks you need to do. Some examples might be: how to reason scientifically, how to identify certain kinds of evidence and how to use evidence to make inferences. Hypothesis testing would be a good example of that.

Self-efficacy comes from the work of Bandura. It basically refers to the extent to which you feel confident that you can perform a task. The idea here is that self-efficacy is crucial. If you have knowledge and you have strategies and other kinds of skills but don’t feel like you can use these skills, it's not going to do you much good. Some unfortunate but very good examples are the issues concerning gender differences in science education. Historically, many women have shied away from the sciences, at least compared to men, for lots of reasons that seem to have mostly to do with efficacy and motivational kinds of concerns rather than ability or knowledge base. So these motivational issues are important for everybody, but particularly for women, in science education settings.

Goals. There is a lot of interesting research done here. The main idea with goals is that the research suggests that a lot of students really aren’t very well equipped to set reasonable goals. They actually need help—scaffolded help—from teachers in setting reasonable goals in science education settings. I’m going to leave it at that for now.

The main idea is simply that self-regulated learning theory tries to look at learning in terms of individual skills students need, as well as motivational components.

Let me just say a last little bit about the idea of situated learning. One of the big ideas among self-regulated learning theorists is that all of these things are accomplished best in real life embedded apprenticeship settings. There’s a real strong movement away from just isolated classroom instruction that’s devoid of the opportunity to get in and do something and interact with peers. The main idea with situated learning, if you’re not familiar with that term, is the chance for students to get in and do a task, like in a lab setting or a field setting, with other kinds of students where there’s plenty of opportunity for students to sort of self-direct one another and provide each other with a lot of dynamic feedback. In a situated learning model the teacher is more of a facilitator, and sort of backs off a little bit from the tip of the pyramid, so to speak.

So again, from my perspective, these three areas are very hot research topics that are being investigated substantially. If you are interested in any of these topics, a very prestigious journal in cognitive psychology that deals with much of this is called *Cognition and Instruction*. It would be available in any major library. *Cognition and Instruction* has far and away the best research in these areas. A lot of the research they publish, and this relates to what I’m going to talk about in just a minute, is in-depth quasi-experimental studies. In my opinion, this is exactly the kind of research that science educators need. They are very well thought out, theoretically planned studies that are conducted in field settings—classrooms or other appropriate field settings. These studies provide tremendous amounts of very rich data to understand how these things play out in real life.

Quantitative Methods

What I want to do now with the rest of my time is quickly overview quantitative methods that would be relevant for science education. I’m going to go through it at a pretty brisk pace because I’m assuming most of you have seen quite a bit of this before. What I’m going to try and encapsulate within the next 10 or 15 minutes is the essence of the research methodologies I teach to graduate students.

What I have tried to do in Figures 1 through 3 is just summarize what the options are. In terms of quantitative research there are three general frameworks or traditions. One is true experiments, another is quasi-experiments and the third is correlational studies. The best way to describe what these are is just by indicating the extent to which they satisfy a number of different constraints, so what I want to do is give you an idea of what the difference is between true and quasi- experiments and then I want to move on to Figure 2, which is what I really want to emphasize. Figure 2 includes what I call foundational criteria. Now these are things I’m guessing you are far less likely to have seen. Foundational criteria are the main benchmarks that educational researchers use to evaluate these kinds of studies.

Research designs

Figure 1

A Comparison of True, Quasi and Correlational Research Designs				
	TE	QE	CORR	Purpose
1. Random Selection	+	-	-	Draw Representative sample from population. [External Validity]
2. Random Assignment	+	-	-	Equivalent Groups. (Eliminate Bias) [Internal Validity]
3. Manipulated Independent Variables	+	+	-	Isolation and control of treatment effect. [Internal Validity]
4. Manipulated Dependent Variables	+	+	-	Select most valid criterion measure. [Internal and Construct Validity]
5. True Control Group	?	-	-	Establish unambiguous baseline needed to detect absolute effect size.
6. Statistical Controls	+	+	+	Reduce within-group variability. [Statistical Conclusion Validity]
7. Pre-tests	+	**$**	+	Provide Measurement Baseline.
Key: **+** = must satisfy; **-** = does not satisfy; **$** = very important; **?** = might satisfy				

Characteristics of the three quantitative research designs are shown in Figure 1. The first one is random selection and the second one is random assignment. What you see in this little summary matrix is a "+" under true experiments. In other words, true experiments are characterized by both random selection and random assignment from some population. Quasi-experiments and correlational studies are not.

Row three refers to having a manipulated independent variable. A manipulated independent variable is the variable you're controlling, the treatment you are interested in testing. True and quasi- experiments both have manipulated independent variables, correlational studies do not.

Both true and quasi-experiments are also characterized by manipulated dependent variables, as seen in row four. This is really crucial because this is what builds experimental control into the study, which in my opinion is really the essence of research.

A true control group (row 5) is some legitimate baseline group that you can use as a measure of what would happen in the absence of a treatment. Ideally you'd like to include these in true experiments but sometimes they might not be possible. Typically in quasi-experiments they're not possible at all because you don't have random assignment, so by definition in quasi-experiments you have non-equivalent groups, which makes it really hard to have a baseline. There's another reason for that, especially if you do research in field settings. What's a real control group? It's the group that gets no instruction. But still the students are doing something. You don't get to just turn them off for three months at a time, so at some level it might be very hard to have a true control group. In lab settings, obviously, it is much easier to have a true control.

The last criteria are a little bit different. Row six is statistical controls, which refers to the opportunity to measure additional variables that you could use, for example, in an analysis of covariance or as some sort of control measure external to the actual plan of study. There is a "+" for each type of study, because that's something you can always tack on. It is really, in my opinion, the only redeeming value of correlational studies. There is also the idea of pretests, listed in the last row. You can include pretest in any study. I have the "$" under quasi-experiments because when I actually teach methods classes one of the things I continually beat into students is that in quasi-experiments, because you don't have equivalent or randomly assigned groups, it is crucial to get pretest information that you can use in some baseline fashion to equate these groups as best you can. Otherwise it is very difficult to make inferences, since you don't necessarily know that the groups are similar.

Let me just make a couple of general comments and then we'll move on again. There's a lot of snobbery, at least in quantitative research, and true experiments are always viewed as sort of the ideal. Correlational studies are viewed as trash, and quasi-experiments are somewhere in the middle. My own personal opinion is I actually think quasi-experiments are probably the ideal study, particularly in terms of physics or science education research, or really any other kind of education research. The reason is that you have the opportunity to build in quite a bit of experimental control, but at the same time you are in a real live setting where you have the opportunity to take your findings and generalize them in a meaningful sort of way. The real problem with true experiments in psychological research is that they are not reflective at all of what goes on in real life.

I'd like to talk about something for a moment. Let's say you do a true experiment or a quasi-experiment. How do you evaluate this stuff? How do you know whether it's a good study or not? Many research methodologists have been interested in this topic lately, and this is one of my major concerns from the standpoint of research. I would say the most comprehensive source for information on these issues is a book by Thomas Cook and Donald Campbell called *Quasi-experimentation: Design & Analysis Issues for Field Settings* (Cook & Campbell, 1979). This book is literally a bible among research methodologists. Any library will have it, and everything that I talk about here is described in gory detail in the first 90 pages. It's a great book for understanding the foundations of quantitative methodology.

Reliability and validity

Figure 2

Reliability
The degree to which an instrument provides a consistent measure of performance.

Internal Validity
The degree to which we can infer a causal relationship between variable X and variable Y.

IV (X) ‡ DV (Y)

[Presumes the elimination of all confounds via random assignment and experimental controls.]

External Validity
The degree to which we can generalize our findings to other settings.

Construct Validity
The degree to which we can measure a construct (e.g., unobservable trait or ability) in a valid manner.

Statistical Conclusion Validity
The degree to which we can trust statistical procedures and results.

I want to talk now about criteria we use to evaluate research. These criteria were actually developed by Cook and Campbell and Campbell and Stanley (this is the same Campbell, but obviously a different Stanley), during the '60s and '70s, and I think they are very crucial. Let me just talk about them quickly. I have five things on Figure 2. One is reliability and then I have four of what I call the foundational criteria. These are: internal validity, external validity, construct validity and statistical conclusion validity. Let me just give you simple examples of these. Again, these are important because the degree to which we satisfy these criteria is the degree to which we conduct good studies, at least according to research methodologists, so I guess the point I want to leave you with is that if you're serious about science education, especially in field settings, you'd be doing yourself a big favor to run down to the library and check this book out. It is pretty expensive—I don't know if you want to buy it, but certainly you ought to read the first 100 pages and get a good sense of this, because the discussions, especially about internal and external validity, are very helpful in terms of framing research and things of that nature.

The idea of reliability is pretty simple. The idea is that whatever kind of instruments we use, and whatever we measure, we should do it consistently. Reliability is simply the idea that our measurements should be very consistent. For educational measurement folks like me, these issues come up constantly. In fact, you may have heard of the Buros Institute of Mental Measurement on campus here. It is affiliated with the Educational Psychology program. It is world famous. Much of the work that's done in the institute looks at the reliability and validity of instruments. It is a sort of a watch dog organization, almost. They talk about these four criteria developed by Cook and Campbell and Campbell and Stanley.

Internal validity is very important. It is the extent to which we can make a causal inference, that an independent variable affects a dependent variable. In my opinion internal validity is really the essence of quantitative research. I assume John Creswell will talk about this a little bit, later. It is not the essence of qualitative research but it is the essence of quantitative research, so what we really need to enhance internal validity is experimental control, random assignment, and things of that nature.

External validity is the degree to which we can generalize our findings to other settings. We do a study and at one level we are interested in internal validity: how well was the study done? Then we become interested in external validity: to what extent can we take these findings and use them to generalize to different settings or populations?

Construct validity is one that educators and psychologists go ballistic about. I’m guessing that in physics education it is generally of less concern. Construct validity is the extent to which we measure the construct of interest. When we talk about constructs we are talking about unobservable phenomena, like love or depression, or we could be talking about self-regulation. What the heck is self-regulation? It's a pretty fuzzy and broad construct. One issue that comes up all the time is the extent to which we actually measure that in a valid way, and clearly that should be an important criterion to use in evaluating our research.

The last one is statistical conclusion validity and I guess what this comes down to is: do we screw up the analysis? I mean, for example, it is entirely possible that we could have data that seriously violates normal distributional properties that most parametric statistics will assume—the t-test and F-test and all those things depend on having normal distributions. We could have an inadequate sample size—many things could go wrong—so we can introduce problems with our research in this domain.

Cook and Campbell identified these four major areas. What really makes the work of Cook and Campbell and Campbell and Stanley great is that they go into a lot of detailed discussion about a variety of threats to these kinds of validity. For example, in internal validity, one threat they discuss is the extent to which poor instrumentation confounds our ability to make this causal inference between the independent and the dependent variable. Their work identifies a lot of major threats to internal validity. When I teach these courses, they are all about introducing these ideas and then spending a lot of time talking about potential threats to internal validity in the context of critiquing real research studies. It is my belief that once you understand what those threats are, you are in pretty good shape. Because once you know what they are and that they are out there, you can circumvent quite a few of them.

Strengths and weaknesses of research designs

Figure 3

A Comparison Of Correlation, Quasi-Experimental, And True-Experimental Designs Across Four Criteria For Validity			
	Correlational	Quasi-Experiment	True-Experiment
Internal Validity	-	+	++
External Validity	+	++	+
Construct Validity	+	++	++
Statistical Conclusion Validity	+	+	+

Ok, we've got these criteria. The thing really we're talking about is contained in this little summary table, Figure 3. I want to offer my opinions in terms of the relative strengths and weaknesses of these three kinds of studies, and maybe make some suggestions for what I think would constitute sort of a reasonable entry point or attack strategy for science education research.

Again we've got the three kinds of studies: correlational, quasi-experimental and true experimental and how they rate in terms of the four foundational criteria I just talked about. For example, the first one, internal validity, which again is our ability to draw a valid causal inference. Correlational studies are very poor at this. Quasi-experiments are better but they still have some problems because often we have non-equivalent control groups. True experiments, because they are very well controlled, usually have very good internal validity. There is a raging debate among psychologists and educators about the relative trade-offs between internal and external validity. Many people are becoming increasingly uncomfortable with true experiments because they are too removed from reality. Maybe John Creswell will talk about this later. There's been a massive change in educational research in the last decade, where some organizations have literally gone from being primarily quantitatively oriented to being qualitatively oriented. A good example, and I'm sure nobody here is familiar with it, is the National Reading Conference. It is a very large organization, and very powerful. Over the last decade they've really done a 180-degree turn. Many of the people now are very qualitative in nature, and they really chastise researchers for during true experiments that are well controlled but ecologically meaningless. There continues to be tremendous debate about these kinds of things.

Let me just say that in many of the review boards I am on in my area, in educational psychology journals, generally about 60% of the work is correlational, about 35% consists of true experiments and about 5% of the papers are quasi-experiments. There is a real paucity of quasi-experiments, which is unfortunate. My personal opinion is that these are the toughest to do and the most informative kinds of studies.

Let me continue to elaborate on Figure 3. Consider external validity. Correlational studies have a little bit. In my opinion true experiments probably have more, but the studies with really great external validity are the quasi-experiments, because they offer not only some degree of experimental control but also the opportunity to collect date in real life settings. Construct validity, again, is the extent to which we accurately measure the construct we’re interested in. Usually, I think, correlational studies are pretty poor. The kind of planning you need to put into quasi- and true experiments usually yields much better construct validity. I say that as a researcher for professional journals. Usually the quasi- and the true experiments are much better. Statistical conclusion validity is the same across all three types of studies, in the sense that you always have the same opportunity to make dumb mistakes anywhere you go.

At this point, let me take a minute or two to offer my opinions after having said all of this from my vantagepoint as an educational researcher. I do a lot of my own research, I review a lot of research for professional journals and I usually teach three or four research methods master’s level and Ph.D. level classes. My opinion is that quasi-experiments are the very best studies you can do, particularly in the context of science learning. Let me tell you why, briefly. They’re the best because they offer you some opportunity to engage in traditional experimental control, manipulating independent variables, and pretests that can be used to establish baselines. When I teach my class, I always like to tell my students that if they do quasi-experiments right they may have some threats to internal validity but they’re not big, massive threats. One can sort of put a put a fence around them. The big payoff, of course, is external validity. I think there’s a tremendous push in educational research to upgrade the external validity of studies, even at the expense of internal validity. Generally I’m very supportive of that. I think it’s a great idea and I’d like to encourage everyone here to do it.

I want to leave you with a suggestion. Maybe you should ask yourself, in terms of your own research, especially physics research with an educational bent, "How important is internal validity versus external validity?" Ultimately you have to make that decision for yourself. If you want to put a little bit more weight on external validity I’d encourage you to get serious about quasi-experiments, and if you still believe that internal validity is the way to go maybe true experiments are right for you. Generally, I think there’s too much correlational research out there already, and we shouldn’t promote any more of it than we have to. That’s my view, anyhow.

<u>References:</u>

Carey, S. (1985). *Conceptual change in childhood.* Cambridge: MIT Press

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & Analysis Issues for Field Settings.* Boston. Houghton Mifflin.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674-689.

The question/answer session for Dr. Schraw's presentation was combined with the one after Dr. Creswell's presentation on qualitative research and appears after the next chapter.