Chapter 1 : Social Research Methods - Knowledge Base - Measurement Validity Types

Validity simply means that a test or instrument is accurately measuring what it's supposed to. Click on the link to visit the individual pages with examples for each type.

Convergent validity Discriminant validity I have to warn you here that I made this list up. First, as mentioned above, I would like to use the term construct validity to be the overarching category. Construct validity is the approximate truth of the conclusion that your operationalization accurately reflects its construct. All of the other terms address this general issue in different ways. Second, I make a distinction between two broad types: In translation validity, you focus on whether the operationalization is a good reflection of the construct. This approach is definitional in nature -- it assumes you have a good detailed definition of the construct and that you can check the operationalization against it. In criterion-related validity, you examine whether the operationalization behaves the way it should given your theory of the construct. This is a more relational approach to construct validity. Translation Validity I just made this one up today! See how easy it is to be a methodologist? I needed a term that described what both face and content validity are getting at. In essence, both of those validity types are attempting to assess the degree to which you accurately translated your construct into the operationalization, and hence the choice of name. Face Validity In face validity, you look at the operationalization and see whether "on its face" it seems like a good translation of the construct. This is probably the weakest way to try to demonstrate construct validity. For instance, you might look at a measure of math ability, read through the questions, and decide that yep, it seems like this is a good measure of math ability i. Or, you might observe a teenage pregnancy prevention program and conclude that, "Yep, this is indeed a teenage pregnancy prevention program. We need to rely on our subjective judgment throughout the research process. We can improve the quality of face validity assessment considerably by making it more systematic. For instance, if you are trying to assess the face validity of a math ability measure, it would be more convincing if you sent the test to a carefully selected sample of experts on math ability testing and they all reported back with the judgment that your measure appears to be a good measure of math ability. Content Validity In content validity, you essentially check the operationalization against the relevant content domain for the construct. For instance, we might lay out all of the criteria that should be met in a program that claims to be a "teenage pregnancy prevention program. Then, armed with these criteria, we could use them as a type of checklist when examining our program. Only programs that meet the criteria can legitimately be defined as "teenage pregnancy prevention programs. But for other constructs e. Criterion-Related Validity In criteria-related validity, you check the performance of your operationalization against some criterion. How is this different from content validity? In content validity, the criteria are the construct definition itself -- it is a direct comparison. In criterion-related validity, we usually make a prediction about how the operationalization will perform based on our theory of the construct. The differences among the different criterion-related validity types is in the criteria they use as the standard for judgment. For instance, we might theorize that a measure of math ability should be able to predict how well a person will do in an engineering-based profession. We could give our measure to experienced engineers and see if there is a high correlation between scores on the measure and their salaries as engineers. A high correlation would provide evidence for predictive validity -- it would show that our measure can correctly predict something that we theoretically think it should be able to predict. For example, if we come up with a way of assessing manic-depression, our measure should be able to distinguish between people who are diagnosed manic-depression and those diagnosed paranoid schizophrenic. If we want to assess the concurrent validity of a new measure of empowerment, we might give the measure to both migrant farm workers and to the farm owners, theorizing that our measure should show that the farm owners are higher in empowerment. As in any discriminating test, the results are more powerful if you are able to show that you can discriminate between two groups that are very similar. Convergent Validity In convergent validity , we examine the degree to which the operationalization is similar to converges on other operationalizations that it theoretically should be similar to. For instance, to show the convergent validity of a Head Start program, we might gather evidence that shows that the program is similar to other

Head Start programs. Or, to show the convergent validity of a test of arithmetic skills, we might correlate the scores on our test with scores on other tests that purport to measure basic math ability, where high correlations would be evidence of convergent validity. Discriminant Validity In discriminant validity , we examine the degree to which the operationalization is not similar to diverges from other operationalizations that it theoretically should be not be similar to. Or, to show the discriminant validity of a test of arithmetic skills, we might correlate the scores on our test with scores on tests that of verbal ability, where low correlations would be evidence of discriminant validity.

Whenever a test or other measuring device is used as part of the data collection process, the validity and reliability of that test is important. Just as we would not use a math test to assess verbal skills, we would not want to use a measuring device for research that was not truly measuring what.

Saul McLeod , published The concept of validity was formulated by Kelly , p. For example a test of intelligence should measure intelligence and not something else such as memory. Internal and External Validity A distinction can be made between internal and external validity. Internal validity refers to whether the effects observed in a study are due to the manipulation of the independent variable and not some other factor. In-other-words there is a causal relationship between the independent and dependent variable. Internal validity can be improved by controlling extraneous variables, using standardized instructions, counter balancing, and eliminating demand characteristics and investigator effects. External validity refers to the extent to which the results of a study can be generalized to other settings ecological validity , other people population validity and over time historical validity. External validity can be improved by setting experiments in a more natural setting and using random sampling to select participants. Assessing the validity of test There there are two main categories of validity used to assess the validity of test i. Face Validity This is the least sophisticated measure of validity. Face validity is simply whether the test appears at face value to measure what it claims to. Accordingly, tests wherein the purpose is unclear have low face validity Nevo, A direct measurement of face validity is obtained by asking people to rate the validity of a test as it appears to them. This rater could use a likert scale to assess face validity. For example, individuals who actually take the test would be well placed to judge its face validity. Also people who work with the test could offer their opinion e. Finally, the researcher could use members of the general public with an interest in the test e. The face validity of a test can be considered a robust construct only if a reasonable level of agreement exists among raters. It should be noted that the term face validity should be avoided when the rating is done by "expert" as content validity is more appropriate. Having face validity does not mean that a test really measures what the researcher intends to measure, but only in the judgment of raters that it appears to do so. Consequently it is a crude and basic measure of validity. However, the implications of items on tests with clear face validity is that they are more vulnerable to social desirability bias. Individuals may manipulate their response to deny or hide problems, or exaggerate behaviors to present a positive images of themselves. It is possible for a test item to lack face validity but still have general validity and measure what it claims to measure. This is good because it reduces demand characteristics and makes it harder for respondents to manipulate their answers. Because most of the original normative sample of the MMPI were good Christians only a depression Christian would think Christ is not coming back. Thus, for this particular religious sample the item does have general validity, but not face validity. Construct Validity Construct validity was invented by Cornball and Meehl This type of validity refers to the extent to which a test captures a specific theoretical construct or trait, and it overlaps with some of the other aspects of validity Construct validity does not concern the simple, factual question of whether a test measures an attribute. To test for construct validity it must be demonstrated that the phenomenon being measured actually exists. So, the construct validity of a test for intelligence, for example, is dependent on a model or theory of intelligence. Construct validity entails demonstrating the power of such a construct to explain a network of research findings and to predict further relationships. However, there is no single method of determining the construct validity of a test. Instead, different methods and approaches are combined to present the overall construct validity of a test. For example, factor analysis and correlational methods can be used. Concurrent validity This is the degree to which a test corresponds to an external criterion that is known concurrently i. If the new test is validated by a comparison with a currently existing criterion, we have concurrent validity. Very often, a new IQ or personality test might be compared with an older but similar test known to have good validity already. Predictive validity This is the degree to which a test accurately predicts a criterion that will occur in the future. For example, a prediction may be made on the basis of a new intelligence test, that high scorers at age 12 will be more likely to obtain university degrees several

years later. If the prediction is born out then the test has predictive validity. Psychological Bulletin, 52, Manual for the Minnesota Multiphasic Personality Inventory. Interpretation of educational measurements. Journal of Educational Measurement, 22 4 , How to reference this article:

Chapter 3 : Test Reliability and Validity Defined

Test Validity versus Experimental Validity Don't confuse this type of validity (often called test validity) with experimental validity, which is composed of internal and external validity. Internal validity indicates how much faith we can have in cause-and-effect statements that come out of our research.

Precision and accuracy Validity [3] of an assessment is the degree to which it measures what it is supposed to measure. This is not the same as reliability , which is the extent to which a measurement gives results that are very consistent. Within validity, the measurement does not always have to be similar, as it does in reliability. However, just because a measure is reliable, it is not necessarily valid. A test cannot be valid unless it is reliable. Validity is also dependent on the measurement measuring what it was designed to measure, and not something else instead. There are many different types of validity. Construct validity Construct validity refers to the extent to which operationalizations of a construct e. It subsumes all other types of validity. For example, the extent to which a test measures intelligence is a question of construct validity. A measure of intelligence presumes, among other things, that the measure is associated with things it should be associated with convergent validity , not associated with things it should not be associated with discriminant validity. Such lines of evidence include statistical analyses of the internal structure of the test including the relationships between responses to different test items. They also include relationships between the test and measures of other constructs. As currently understood, construct validity is not distinct from the support for the substantive theory of the construct that the test is designed to measure. As such, experiments designed to reveal aspects of the causal role of the construct also contribute to constructing validity evidence. For example, does an IQ questionnaire have items covering all areas of intelligence discussed in the scientific literature? Content validity evidence involves the degree to which the content of the test matches a content domain associated with the construct. For example, a test of the ability to add two numbers should include a range of combinations of digits. A test with only one-digit numbers, or only even numbers, would not have good coverage of the content domain. Content related evidence typically involves a subject matter expert SME evaluating test items against the test specifications. Items are chosen so that they comply with the test specification which is drawn up through a thorough examination of the subject domain. The experts will be able to review the items and comment on whether the items cover a representative sample of the behavior domain. Face validity[edit] Face validity is an estimate of whether a test appears to measure a certain criterion; it does not guarantee that the test actually measures phenomena in that domain. Measures may have high validity, but when the test does not appear to be measuring what it is, it has low face validity. Indeed, when a test is subject to faking malingering , low face validity might make the test more valid. Considering one may get more honest answers with lower face validity, it is sometimes important to make it appear as though there is low face validity whilst administering the measures. Face validity is very closely related to content validity. While content validity depends on a theoretical basis for assuming if a test is assessing all domains of a certain criterion e. To answer this you have to know, what different kinds of arithmetic skills mathematical skills include face validity relates to whether a test appears to be a good measure or not. This judgment is made on the "face" of the test, thus it can also be judged by the amateur. Face validity is a starting point, but should never be assumed to be probably valid for any given purpose, as the "experts" have been wrong beforeâ€"the Malleus Malificarum Hammer of Witches had no support for its conclusions other than the self-imagined competence of two "experts" in "witchcraft detection," yet it was used as a "test" to condemn and burn at the stake tens of thousands men and women as "witches. In other words, it compares the test with other measures or outcomes the criteria already held to be valid. For example, employee selection tests are often validated against measures of job performance the criterion , and IQ tests are often validated against measures of academic performance the criterion. If the test data and criterion data are collected at the same time, this is referred to as concurrent validity evidence. If the test data are collected first in order to predict criterion data collected at a later point in time, then this is referred to as predictive validity evidence. Concurrent validity[edit] Concurrent validity refers to the degree to which the operationalization correlates

with other measures of the same construct that are measured at the same time. When the measure is compared to another measure of the same type, they will be related or correlated. Returning to the selection test example, this would mean that the tests are administered to current employees and then correlated with their scores on performance reviews. Predictive validity[edit] Predictive validity refers to the degree to which the operationalization can predict or correlate with other measures of the same construct that are measured at some time in the future. Again, with the selection test example, this would mean that the tests are administered to applicants, all applicants are hired, their performance is reviewed at a later time, and then their scores on the two measures are correlated. This is also when measurement predicts a relationship between what is measured and something else; predicting whether or not the other thing will happen in the future. High correlation between ex-ante predicted and ex-post actual outcomes is the strongest proof of validity. Experimental validity[edit] The validity of the design of experimental research studies is a fundamental part of the scientific method , and a concern of research ethics. Without a valid design, valid scientific conclusions cannot be drawn. Internal validity[edit] Internal validity is an inductive estimate of the degree to which conclusions about causal relationships can be made e. Good experimental techniques, in which the effect of an independent variable on a dependent variable is studied under highly controlled conditions, usually allow for higher degrees of internal validity than, for example, single-case designs. Eight kinds of confounding variable can interfere with internal validity i. History, the specific events occurring between the first and second measurements in addition to the experimental variables Maturation, processes within the participants as a function of the passage of time not specific to particular events , e. Testing, the effects of taking a test upon the scores of a second testing. Instrumentation, changes in calibration of a measurement tool or changes in the observers or scorers may produce changes in the obtained measurements. Statistical regression, operating where groups have been selected on the basis of their extreme scores. Selection, biases resulting from differential selection of respondents for the comparison groups. Experimental mortality, or differential loss of respondents from the comparison groups. In other words, it is about whether findings can be validly generalized. If the same research study was conducted in those other cases, would it get the same results? A major factor in this is whether the study sample e. Other factors jeopardizing external validity are: Reactive or interaction effect of testing, a pretest might increase the scores on a posttest Interaction effects of selection biases and the experimental variable. Reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in non-experimental settings Multiple-treatment interference, where effects of earlier treatments are not erasable. Ecological validity[edit] Ecological validity is the extent to which research results can be applied to real-life situations outside of research settings. To be ecologically valid, the methods, materials and setting of a study must approximate the real-life situation that is under investigation. Ecological validity is partly related to the issue of experiment versus observation. Typically in science, there are two domains of research: The purpose of experimental designs is to test causality, so that you can infer A causes B or B causes A. Then you can still do research, but it is not causal, it is correlational. You can only conclude that A occurs together with B. Both techniques have their strengths and weaknesses. Relationship to internal validity[edit] On first glance, internal and external validity seem to contradict each other â€" to get an experimental design you have to control for all interfering variables. That is why you often conduct your experiment in a laboratory setting. While gaining internal validity excluding interfering variables by keeping them constant you lose ecological or external validity because you establish an artificial laboratory setting. On the other hand, with observational research you can not control for interfering variables low internal validity but you can measure in the natural ecological environment, at the place where behavior normally occurs. However, in doing so, you sacrifice internal validity. The apparent contradiction of internal validity and external validity is, however, only superficial. The question of whether results from a particular study generalize to other people, places or times arises only when one follows an inductivist research strategy. If the goal of a study is to deductively test a theory, one is only concerned with factors which might undermine the rigor of the study, i. Diagnostic validity[edit] In psychiatry there is a particular issue with assessing the validity of the diagnostic categories themselves. Robins and Guze proposed in what were to become influential formal criteria for establishing the

validity of psychiatric diagnoses. They listed five criteria: Kendler in distinguished between: On this basis, he argues that a Robins and Guze criterion of "runs in the family" is inadequately specific because most human psychological and physical traits would qualify - for example, an arbitrary syndrome comprising a mixture of "height over 6 ft, red hair, and a large nose" will be found to "run in families" and be " hereditary ", but this should not be considered evidence that it is a disorder. Kendler has further suggested that " essentialist " gene models of psychiatric disorders, and the hope that we will be able to validate categorical psychiatric diagnoses by "carving nature at its joints" solely as a result of gene discovery, are implausible. Perri and Lichtenwald provide a starting point for a discussion about a wide range of reliability and validity topics in their analysis of a wrongful murder conviction.

Chapter 4 : "Understanding Reliability and Validity in Qualitative Research" by Nahid Golafshani

Test-retest reliability is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time.

Validity refers to how well a test measures what it is purported to measure. Why is it necessary? While reliability is necessary, it alone is not sufficient. For a test to be reliable, it also needs to be valid. For example, if your scale is off by 5 lbs, it reads your weight every day with an excess of 5lbs. The scale is reliable because it consistently reports the same weight every day, but it is not valid because it adds 5lbs to your true weight. It is not a valid measure of your weight. Types of Validity 1. Face Validity ascertains that the measure appears to be assessing the intended construct under study. The stakeholders can easily assess face validity. If the stakeholders do not believe the measure is an accurate assessment of the ability, they may become disengaged with the task. If a measure of art appreciation is created all of the items should be related to the different components and types of art. If the questions are regarding historical time periods, with no reference to any artistic movement, stakeholders may not be motivated to give their best effort or invest in this measure because they do not believe it is a true assessment of art appreciation. Construct Validity is used to ensure that the measure is actually measure what it is intended to measure i. The experts can examine the items and decide what that specific item is intended to measure. Students can be involved in this process to obtain their feedback. The questions are written with complicated wording and phrasing. It is important that the measure is actually assessing the intended construct, rather than an extraneous factor. Criterion-Related Validity is used to predict future or current performance - it correlates test results with another criterion of interest. If a physics program designed a measure to assess cumulative student learning throughout the major. The new measure could be correlated with a standardized measure of ability in this discipline, such as an ETS field test or the GRE subject test. The higher the correlation between the established measure and new measure, the more faith stakeholders can have in the new assessment tool. Formative Validity when applied to outcomes assessment it is used to assess how well a measure is able to provide information to help improve the program under study. If the measure can provide information that students are lacking knowledge in a certain area, for instance the Civil Rights Movement, then that assessment tool is providing meaningful information that can be used to improve the course or program requirements. Sampling Validity similar to content validity ensures that the measure covers the broad range of areas within the concept under study. Not everything can be covered, so items need to be sampled from all of the domains. When designing an assessment of learning in the theatre department, it would not be sufficient to only cover issues related to acting. Other areas of theatre such as lighting, sound, functions of stage managers should all be included. The assessment should reflect the content area in its entirety. What are some ways to improve validity? Make sure your goals and objectives are clearly defined and operationalized. Expectations of students should be written down. Match your assessment measure to your goals and objectives. Additionally, have the test reviewed by faculty at other schools to obtain feedback from an outside party who is less invested in the instrument. Get students involved; have the students look over the assessment for troublesome wording, or other difficulties. If possible, compare your measure with other measures, or data that may be available. Standards for educational and psychological testing. Methods in Behavioral Research 7th ed. Educational Measurement 2nd ed. American Council on Education. The Center for the Enhancement of Teaching. How to improve test reliability and validity:

Chapter 5 : What is Validity in Psychology | Simply Psychology

Research validity in surveys relates to the extent at which the survey measures right elements that need to be measured. In simple terms, validity refers to how well an instrument as measures what it is intended to measure.

We might say that a measure is a valid one, or that a valid sample was drawn, or that the design had strong validity. But all of those statements are technically incorrect. Technically, we should say that a measure leads to valid conclusions or that a sample enables valid inferences, and so on. We make lots of different inferences or conclusions while conducting research. Many of these are related to the process of doing research and are not the major hypotheses of the study. Nevertheless, like the bricks that go into building a wall, these intermediate process and methodological propositions provide the foundation for the substantive conclusions that we wish to address. For instance, virtually all social research involves measurement or observation. And, whenever we measure or observe we are concerned with whether we are measuring what we intend to measure or with how our observations are influenced by the circumstances in which they are made. We reach conclusions about the quality of our measures -- conclusions that will play an important role in addressing the broader substantive issues of our study. When we talk about the validity of research, we are often referring to these to the many conclusions we reach about the quality of different parts of our research methodology. We subdivide validity into four types. Each type addresses a specific methodological question. In order to understand the types of validity, you have to know something about how we investigate a research question. Because all four validity types are really only operative when studying causal questions, we will use a causal study to set the context. The figure shows that there are really two realms that are involved in research. The first, on the top, is the land of theory. It is what goes on inside our heads as researchers. It is where we keep our theories about how the world operates. The second, on the bottom, is the land of observations. It is the real world into which we translate our ideas -- our programs, treatments, measures and observations. When we conduct research, we are continually flitting back and forth between these two realms, between what we think about the world and what is going on in it. When we are investigating a cause-effect relationship, we have a theory implicit or otherwise of what the cause is the cause construct. For instance, if we are testing a new educational program, we have an idea of what it would look like ideally. Similarly, on the effect side, we have an idea of what we are ideally trying to affect and measure the effect construct. But each of these, the cause and the effect, has to be translated into real things, into a program or treatment and a measure or observational method. We use the term operationalization to describe the act of translating a construct into its manifestation. In effect, we take our idea and describe it as a series of operations or procedures. Now, instead of it only being an idea in our minds, it becomes a public entity that anyone can look at and examine for themselves. It is one thing, for instance, for you to say that you would like to measure self-esteem a construct. But when you show a ten-item paper-and-pencil self-esteem measure that you developed for that purpose, others can look at it and understand more clearly what you intend by the term self-esteem. Now, back to explaining the four validity types. They build on one another, with two of them conclusion and internal referring to the land of observation on the bottom of the figure, one of them construct emphasizing the linkages between the bottom and the top, and the last external being primarily concerned about the range of our theory on the top. Assume that we took these two constructs, the cause construct the WWW site and the effect understanding , and operationalized them -- turned them into realities by constructing the WWW site and a measure of knowledge of the course material. Here are the four validity types and the question each addresses: In this study, is there a relationship between the two variables? There are several conclusions or inferences we might draw to answer such a question. We could, for example, conclude that there is a relationship. We might conclude that there is a positive relationship. We might infer that there is no relationship. We can assess the conclusion validity of each of these conclusions or inferences. Assuming that there is a relationship in this study, is the relationship a causal one? Both could, for example, be caused by the same factor. For instance, it may be that wealthier students who have greater resources would be more likely to use have access to a WWW site and would excel on objective tests. When we want to make a claim that our program or treatment caused the outcomes in our

study, we can consider the internal validity of our causal claim. Assuming that there is a causal relationship in this study, can we claim that the program reflected well our construct of the program and that our measure reflected well our idea of the construct of the measure? In simpler terms, did we implement the program we intended to implement and did we measure the outcome we wanted to measure? In yet other terms, did we operationalize well the ideas of the cause and the effect? When our research is over, we would like to be able to conclude that we did a credible job of operationalizing our constructs -- we can assess the construct validity of this conclusion. Assuming that there is a causal relationship in this study between the constructs of the cause and the effect, can we generalize this effect to other persons, places or times? We are likely to make some claims that our research findings have implications for other groups and individuals in other settings and at other times. When we do, we can examine the external validity of these claims. Notice how the question that each validity type addresses presupposes an affirmative answer to the previous one. This is what we mean when we say that the validity types build on one another. The figure shows the idea of cumulativeness as a staircase, along with the key question for each validity type. For any inference or conclusion, there are always possible threats to validity -- reasons the conclusion or inference might be wrong. Ideally, one tries to reduce the plausibility of the most likely threats to validity, thereby leaving as most plausible the conclusion reached in the study. For instance, imagine a study examining whether there is a relationship between the amount of training in a specific technology and subsequent rates of use of that technology. Because the interest is in a relationship, it is considered an issue of conclusion validity. Assume that the study is completed and no significant correlation between amount of training and adoption rates is found. On this basis it is concluded that there is no relationship between the two. How could this conclusion be wrong -- that is, what are the "threats to validity"? Perhaps the sample size is too small or the measure of amount of training is unreliable. Or maybe assumptions of the correlational test are violated given the variables used. Perhaps there were random irrelevancies in the study setting or random heterogeneity in the respondents that increased the variability in the data and made it harder to see the relationship of interest. The inference that there is no relationship will be stronger -- have greater conclusion validity -- if one can show that these alternative explanations are not credible. The distributions might be examined to see if they conform with assumptions of the statistical test, or analyses conducted to determine whether there is sufficient statistical power. The theory of validity, and the many lists of specific threats, provide a useful scheme for assessing the quality of research conclusions. The theory is general in scope and applicability, well-articulated in its philosophical suppositions, and virtually impossible to explain adequately in a few minutes. As a framework for judging the quality of evaluations it is indispensable and well worth understanding.

Chapter 6 : Reliability and Validity

Again, high test-retest correlations make sense when the construct being measured is assumed to be consistent over time, which is the case for intelligence, self-esteem, and the Big Five personality dimensions.

Standard Validity is described as the degree to which a research study measures what it intends to measure. There are two main types of validity, internal and external. Internal validity refers to the validity of the measurement and test itself, whereas external validity refers to the ability to generalise the findings to the target population. Both are very important in analysing the appropriateness, meaningfulness and usefulness of a research study. However, here I will focus on the validity of the measurement technique i. The 4 main types of validity There are 4 main types of validity used when assessing internal validity. Each type views validity from a different perspective and evaluates different relationships between measurements. Face validity-This refers to whether a technique looks as if it should measure the variable it intends to measure. For example, a method where a participant is required to click a button as soon as a stimulus appears and this time is measured appears to have face validity for measuring reaction time. An example of analysing research for face validity by Hardesty and Bearden can be found here. Concurrent validity-This compares the results from a new measurement technique to those of a more established technique that claims to measure the same variable to see if they are related. Often two measurements will behave in the same way, but are not necessarily measuring the same variable, therefore this kind of validity must be examined thoroughly. Predictive validity-This is when the results obtained from measuring a construct can be accurately used to predict behaviour. There are obvious limitations to this as behaviour cannot be fully predicted to great depths, but this validity helps predict basic trends to a certain degree. A meta-analysis by van IJzendoorn examines the predictive validity of the Adult Attachment Interview. Construct validity-This is whether the measurements of a variable in a study behave in exactly the same way as the variable itself. This involves examining past research regarding different aspects of the same variable. For example, using measurements of weight to measure the variable height has concurrent validity as weight generally increases as height increases, however it lacks construct validity as weight fluctuates based on food deprivation whereas height does not. What are the threats to Internal Validity? Factors that can effect internal validity can come in many forms, and it is important that these are controlled for as much as possible during research to reduce their impact on validity. The term history refers to effects that are not related to the treatment that may result in a change of performance over time. Instrumental bias refers to a change in the measuring instrument over time which may change the results. This is often evident in behavioural observations where the practice and experience of the experimenter influences their ability to notice certain things and changes their standards. A main threat to internal validity is testing effects. Often participants can become tired or bored during an experiment, and previous tests may influence their performance. This is often counterbalanced in experimental studies so that participants receive the tasks in a different order to reduce their impact on validity. So why is validity important? If the results of a study are not deemed to be valid then they are meaningless to our study. If it does not measure what we want it to measure then the results cannot be used to answer the research question, which is the main aim of the study. These results cannot then be used to generalise any findings and become a waste of time and effort. It is important to remember that just because a study is valid in one instance it does not mean that it is valid for measuring something else. Reliability is the consistency of results when the experiment is replicated under the same conditions, which is very different to validity. These two evaluations of research studies are independent factors, therefore a study can be reliable without being valid, and vice versa, as demonstrated here this resource also provides more information on types of validity and threats. However, a good study will be both reliable and valid. So to conclude, validity is very important in a research study to ensure that our results can be used effectively, and variables that may threaten validity should be controlled as much as possible.

Chapter 7 : Test validity - Wikipedia

Test validity incorporates a number of different validity types, including criterion validity, content validity and construct validity. If a research project scores highly in these areas, then the overall test validity is high.

Establishing Validity in Qualitative Research Establishing Validity in Qualitative Research The following module discusses reliability and validity in qualitative research, with an emphasis on establishing credibility and transferability. Define and reliability and validity in qualitative research. Discuss the importance of establishing validity. List strategies used by researchers to improve reliability and validity. Qualitative research is based on subjective, interpretive and contextual data, making the findings are more likely to be scrutinized and questioned. Therefore, it is critical that researchers take steps to ensure the reliability and validity of their research findings. The findings must be believable, consistent, applicable and credible if they are to be useful to readers and other researchers. Reliability refers to consistency with which the research will produce the same results if repeated. Validity refers to accuracy or correctness of the findings. The following video provides an excellent introductory overview to reliability and validity, including an explanation of terms and specific examples. Qualitative research has become increasingly popular in the past two decades. Therefore, much time has been spent reviewing ways to judge the reliability and validity of qualitative research findings. In order to withstand the scrutiny, researchers should spend time giving serious consideration to the following four aspects: Credibility - Often called internal validity, refers to the believability and trustworthiness of the findings. This depends more on the richness of the data gathered than on the quantity of data. The participants of the study are the only ones that decide if the results actually reflect the phenomena being studied and therefore, it is important that participants feel the findings are credible and accurate. Triangulation is a commonly used method for verifying accuracy that involves cross-checking information from multiple perspectives. The link in Resources Links on the left describes different types of triangulation methods. Transferability - Often called external validity, refers to the degree that the findings of the research can be transferred to other contexts by the readers. This means that the results are generalizable and can be applied to other similar settings, populations, situations and so forth. Researchers should thoroughly describe the context of the research to assist the reader in being able to generalize the findings and apply them appropriately. Dependability - Otherwise known as reliability, refers to the consistency with which the results could be repeated and result in similar findings. The dependability of the findings also lends legitimacy to the research method. Because the nature of qualitative research often results in an ever changing research setting and changing contexts, it is important that researcher document all aspects of any changes or unexpected occurrences to further explain the findings. This is also important for other researchers who may want to replicate the study. Confirmability - A measure of the objectivity used in evaluating the results, describes how well the research findings are supported by the actual data collected when examined by other researchers. Researchers bring their own unique perspectives to the research process and data interpretation can be somewhat subjective in qualitative research. If findings are corroborated or confirmed by others who examine the data, then no inappropriate biases impacted the data analysis. Criteria for assessing interpretive validity in qualitative research. Issues of validity in qualitative research. Reliability and validity in qualitative research. Validity and qualitative research: A practical guide to research methods. Validity in qualitative research. Qualitative health research, 11 4 ,

Chapter 8 : What is validity and why is it important in research? | psucd8

Like reliability and validity as used in quantitative research are providing springboard to examine what these two terms mean in the qualitative research paradigm, triangulation as used in quantitative research to test the reliability and validity can also illuminate some ways to test or maximize the validity and reliability of a qualitative study.

Example usability problems include: Students are asked to rate a lesson immediately after class, but there are only a few minutes before the next class begins problem with administration. Students are asked to keep self-checklists of their after school activities, but the directions are complicated and the item descriptions confusing problem with interpretation. Validity and reliability concerns discussed below will help alleviate usability issues. For now, we can identify five usability considerations: How long will it take to administer? Are the directions clear? How easy is it to score? Do equivalent forms exist? Have any problems been reported by others who used it? It is best to use an existing instrument, one that has been developed and tested numerous times, such as can be found in the Mental Measurements Yearbook. We will turn to why next. Validity Validity is the extent to which an instrument measures what it is supposed to measure and performs as it is designed to perform. As a process, validation involves collecting and analyzing data to assess the accuracy of an instrument. There are numerous statistical tests and measures to assess the validity of quantitative instruments, which generally involves pilot testing. The remainder of this discussion focuses on external validity and content validity. External validity is the extent to which the results of a study can be generalized from a sample to a population. Establishing eternal validity for an instrument, then, follows directly from sampling. Recall that a sample should be an accurate representation of a population, because the total population may not be available. An instrument that is externally valid helps obtain population generalizability, or the degree to which a sample represents the population. Content validity refers to the appropriateness of the content of an instrument. In other words, do the measures questions, observation logs, etc. This is particularly important with achievement tests. Consider that a test developer wants to maximize the validity of a unit test for 7th grade mathematics. This would involve taking representative questions from each of the sections of the unit and evaluating them against the desired outcomes. Reliability Reliability can be thought of as consistency. Does the instrument consistently measure what it is intended to measure? It is not possible to calculate reliability; however, there are four general estimators that you may encounter in reading research: The consistency of a measure evaluated over time. The reliability of two tests constructed the same way, from the same content. Relating Reliability and Validity Reliability is directly related to the validity of the measure. There are several important principles. First, a test can be considered reliable, but not valid. Consider the SAT, used as a predictor of success in college. It is a reliable test high scores relate to high GPA , though only a moderately valid indicator of success due to the lack of structured environment â€" class attendance, parent-regulated study, and sleeping habits â€" each holistically related to success. Second, validity is more important than reliability. Using the above example, college admissions may consider the SAT a reliable test, but not necessarily a valid measure of other quantities colleges seek, such as leadership capability, altruism, and civic involvement. Finally, the most useful instrument is both valid and reliable. Proponents of the SAT argue that it is both. Validity and Reliability in Qualitative Research Thus far, we have discussed Instrumentation as related to mostly quantitative measurement. Some qualitative researchers reject the concept of validity due to the constructivist viewpoint that reality is unique to the individual, and cannot be generalized. These researchers argue for a different standard for judging research quality.

Chapter 9 : Validity (statistics) - Wikipedia

Validity encompasses the entire experimental concept and establishes whether the results obtained meet all of the requirements of the scientific research method. For example, there must have been randomization of the sample groups and appropriate care and diligence shown in the allocation of controls.

Researchers who conduct scientific studies are often motivated by external factors, such as the desire to get published, advance their careers, receive funding, or seek certain results. As a consequence, a significant number of scientific studies are biased and unreliable. Guess whose findings your doctor ends up reading about in the journal, and you end up hearing about on the evening news? Reliable studies use random samples whenever possible, utilize appropriate sample sizes, avoid biases, and should be conducted by researchers who are not influenced by funding or the desire to seek certain results. Randomization Randomization in studies is critical to ensuring the validity of research. Randomized trials in the clinical setting generally assign groups of randomly chosen individuals to either receive a treatment or to receive a placebo or no treatment. Participation in each group is determined randomly using a computer or random number generator before the trials begin in order to ensure that there is no systematic bias in either group. In this way, the two groups are as similar as possible at the start of the study. At the end of the study, if one group has a better outcome than the other, the investigators will be able to conclude with some confidence that one intervention is better than the other. The wording of questions, how studies are designed, which measurements are chosen for analysis, and how results are presented can all influence the validity of a study. Sample Size In the medical setting, research is done to find a solution to a particular problem, or to assess the impact of a treatment. In an ideal situation, the entire desired population should be studied in order to reach a conclusion. However, surveying or carrying out a study with an entire population is almost impossible and very costly. Thus, a sample representative of the population is used, and the data is analyzed and then conclusions are drawn and extrapolated to the population under study. It is important to have an appropriately sized sample to achieve reliable results and high statistical power â€" the ability to discern a difference between study groups when a difference truly exists. An insufficient sample size is more likely to produce false negatives and inconsistent results. On the other hand, too large of a sample is not recommended because it can be unwieldy to manage, and it is a waste of time and money if an answer can be accurately found from a smaller sample. The causes of bias can be related to the manner in which study subjects are chosen, the method in which study variables are collected or measured, the attitudes or preferences of an investigator, and the lack of control of confounding variablesâ€¦ in epidemiologic terms bias can lead to incorrect estimates of association, or, more simply, the observed study results will tend to be in error and different from the true results. Although bias in research can never be completely eliminated, it can be drastically reduced by carefully considering factors that have the potential to influence results during both the design and analysis phases of a study. The most common types of bias in research studies are selection biases, measurement biases and intervention biases. Selection bias also may occur if a study compares a treatment and control group, but they are inherently different. If selection bias is present in a study, it is likely to influence the outcome and conclusions of the study. This can occur due to leading questions, which in some way unduly favor one response over another, or measurement bias may be due to social desirability and the fact that most people like to present themselves in a favorable light, and therefore, will not respond honestly. The Importance of Having a Baseline For Comparison When assessing behavioral changes, it is essential to have a baseline or control group for comparison. It is important to evaluate the impact of a program and determine whether the program actually had an impact, or if what happened would have occurred regardless of the implementation of the program. Such an evaluation could yield useful information to program implementers. But it could not be considered a rigorous evaluation of the effects of the program if there are good reasons to believe that scores might have changed even without the program. For example, many programs and organizations have developed in recent years to make cell phones and mobile technology available to rural areas. A program evaluation could report that a program was able to increase cell phone ownership in a village over a 3-year period. However, conclusive impact cannot be

discerned if the statistics compare the village before and after the program intervention. A credible comparison group is important to determine or prove the full impact of a program or intervention. For example, one could locate other rural areas in the same country that had cell phone usage rates that were comparable to those of the intervention site prior to the start of the program. After the same specified time period i. This type of comparison group is essential when evaluating behavioral changes. Footnotes 1 Freedman, D. Accessed on 17 November