Continue

Examples of valid

Research validity is a critical concept in academic research, referring to the degree to which a study accurately measures what it intends to measure. It assesses the credibility of research results, determining whether the conclusions drawn from a study genuinely reflect the reality of the situation or variables being studied. Without validity, research findings lose their utility and relevance, making it essential for scholars to ensure that their studies are both valid and reliable. Research validity can be broadly divided into two main categories: Internal Validity and External Validity. Each category has distinct subtypes that serve to address specific areas of concern in the research design and findings. Internal validity focuses on the extent to which a study's results can be attributed to the variables manipulated or measured rather than to other extraneous factors. High internal validity means that the study effectively demonstrates causation between the variables being examined. Construct Validity: This subtype of internal validity ensures that the operational definitions of variables accurately represent the theoretical constructs. For instance, if a study aims to measure intelligence, construct validity assesses whether the tools used (like IQ tests) truly reflect intelligence rather than other factors, such as memory or educational background. Content Validity: Content validity is concerned with whether the measurement covers the entire scope of the concept being studied. For instance, in a study on job satisfaction, content validity would ensure that questions cover all relevant aspects of job satisfaction, such as work environment, compensation, and career development. Criterion Validity: Criterion validity evaluates how well one measure predicts an outcome based on another established measure. It is divided into two types: concurrent validity (assessed when the measurements are taken at the same time) and predictive validity (assessed when the measure predicts a future outcome). For example, if a new test of college readiness correlates strongly with existing SAT scores, it would demonstrate high criterion validity. External validity refers to the extent to which the findings of a study can be generalized to other populations, settings, or times. High external validity indicates that the results are applicable beyond the specific conditions of the study. Population Validity: This type of validity is concerned with the extent to which the findings can be generalized to larger populations. If a study on employee motivation is conducted only among tech industry workers, population validity examines whether the findings would apply to employees in other sectors. Ecological Validity: Ecological validity focuses on whether the findings can be generalized to real-world settings. For example, if a study on classroom behavior is conducted in a highly controlled lab setting, its ecological validity would be questionable because it may not accurately reflect behavior in an actual classroom environment. Various factors can threaten the validity of research, potentially leading to biased or inaccurate conclusions. Understanding these threats is essential for researchers to mitigate their impact. Selection Bias: Differences in the characteristics of participants across groups can distort study results. Maturation: Participants may change over time due to natural developmental processes, affecting study outcomes. History: External events occurring during the study period may influence results, particularly in longitudinal studies. Instrumentation: Changes in measurement tools or procedures during the study can introduce variability unrelated to the variables being studied. Sample Characteristics: Using a non-representative sample limits the generalizability of the findings. Setting Effects: Conducting research in an artificial environment can affect behavior, reducing ecological validity. Timing of Measurement: Results collected at one time may not be applicable at another, especially if societal or cultural factors shift. To increase the validity of their research, scholars can implement various strategies: Randomization: Randomly assigning participants to different groups helps control for selection bias and improves internal validity. Control Groups: Using a control group allows researchers to account for extraneous variables, providing a comparison to the experimental group. Blinding: Blinding participants (and, when possible, researchers) to the conditions helps reduce bias in responses and analysis. Pilot Testing: Conducting pilot tests allows researchers to refine their measurement tools, improving construct and content validity. Replication: Replicating a study in different contexts or with different populations enhances external validity, demonstrating the robustness of the findings. Let's say a researcher wants to study how the amount of sleep high school students get each night affects their academic performance. Internal Validity Example Construct Validity: To measure academic performance, the researcher decides to use students' GPAs. Construct validity examines whether GPA accurately represents academic success. If the study only looks at GPA, it might miss other aspects of academic performance, such as critical thinking or classroom engagement. To strengthen construct validity, the researcher could include additional measures, like standardized test scores or teacher evaluations, to capture the full scope of academic performance. Content Validity: The researcher includes a survey that asks about both weekday and weekend sleep patterns. This ensures that the study covers all relevant aspects of sleep behavior and isn't limited to only one part of the week, enhancing content validity. External Validity Example Population Validity: Suppose this study only involves students from a single high school in a wealthy district. While the findings may apply to similar schools, they may not be generalizable to students in rural or low-income areas, who may experience different sleep-related factors (e.g., longer commute times or part-time jobs). To improve population validity, the researcher could include students from various backgrounds and locations. Ecological Validity: The researcher collects data through a lab experiment, where students sleep in a controlled environment with specific lights and sounds. This setup could limit ecological validity, as it doesn't fully mimic a real bedroom environment. To improve ecological validity, the researcher could collect data from students' actual homes to observe sleep patterns in a natural setting. Strategies to Improve Validity in This Study Random Assignment: The researcher could randomly assign students to get different amounts of sleep to ensure that any differences in academic performance are likely due to sleep rather than other factors. Control Group: Including a control group that maintains a regular sleep schedule allows for comparisons and helps control for extraneous factors that might affect performance. Blinding: If possible, blinding both students and teachers to the study conditions can reduce any influence of expectations on behavior or performance. By addressing these aspects, the researcher can increase both the internal and external validity of the study, making the findings more accurate and generalizable. Validity is a foundational aspect of rigorous research, underpinning the credibility and applicability of a study's findings. By understanding and addressing various types of validity and the threats associated with them, researchers can design studies that provide reliable and generalizable insights. Whether assessing internal or external validity, it is crucial for scholars to implement strategies that strengthen their research, ensuring that the results contribute meaningfully to the body of knowledge. Campbell, D. T., & Stanley, J. C. (1963). Experimental and Quasi-Experimental Designs for Research. Houghton Mifflin. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin. Trochim, W. M., Donnelly, J. P., & Arora, K. (2015). Research Methods: The Essential Knowledge Base. Cengage Learning. Babbie, E. R. (2020). The Practice of Social Research. Cengage Learning. Validity refers to whether or not a test or an experiment is actually doing what it is intended to do. Validity sits upon a spectrum. For example: Low Validity: Most people now know that the standard IQ test does not actually measure intelligence or predict success in life. High Validity: By contrast, a standard pregnancy test is about 99% accurate, meaning it has very high validity and is therefore a very reliable test. There are many ways to determine validity. Most of them are defined below. Face validity refers to whether a scale "appears" to measure what it is supposed to measure. That is, do the questions seem to be logically related to the construct under study. For example, a personality scale that measures emotional intelligence should have questions about self-awareness and empathy. It should not have questions about math or chemistry. One common way to assess face validity is to ask a panel of experts to examine the scale and rate it's appropriateness as a tool for measuring the construct. If the experts agree that the scale measures what it has been designed to measure, then the scale is said to have face validity. If a scale, or a test, doesn't have face validity, then people taking it won't be serious. Conbach explains it in the following way: "When a patient loses faith in the medicine his doctor prescribes, it loses much of its power to improve his health. He may skip doses, and in the end may decide doctors cannot help him and let treatment lapse all together. For similar reasons, when selecting a test one must consider how worthwhile it will appear to the participant who takes it and other laymen who will see the results" (Cronbach, 1970, p. 182). Content validity refers to whether a test or scale is measuring all of the components of a given construct. For example, if there are five dimensions of emotional intelligence (EQ), then a scale that measures EQ should contain questions regarding each dimension. Similar to face validity, content validity can be assessed by asking subject matter experts (SMEs) to examine the test. If experts agree that the test includes items that assess every domain of the construct, then the test has content validity. For example, the math portion of the SAT contains questions that require skills in many types of math: arithmetic, algebra, geometry, calculus, and many others. Since there are questions that assess each type of math, then the test has content validity. The developer of the test could ask SMEs to rate the test's construct validity. If the SMEs all give the test high ratings, then it has construct validity. Construct validity is the extent to which a measurement tool is truly assessing what it has been designed to assess. There are two main methods of assessing construct validity: convergent and discriminant validity. Convergent validity involves taking two tests that are supposed to measure the same construct and administering them to a sample of participants. The higher the correlation between the two tests, the stronger the construct validity. With divergent validity, two tests that measure completely different constructs are administered to the same sample of participants. Since the tests are measuring different constructs, there should be a very low correlation between the two. Internal validity refers to whether or not the results of an experiment are due to the manipulation of the independent, or treatment, variables. For example, a researcher wants to examine how temperature affects willingness to help, so they have research participants wait in a room. There are different rooms, one has the temperature set at normal, one at moderately warm, and the other at very warm. During the next phase of the study, participants are asked to donate to a local charity before taking part in the rest of the study. The results showed that as the temperature of the room increased, donations decreased. On the surface, it seems as though the study has internal validity: room temperature affected donations. However, even though the experiment involved three different rooms set at different temperatures, each room was a different size. The smallest room was the warmest and the normal temperature room was the largest. Now, we don't know if the donations were affected by room temperature or room size. So, the study has questionable internal validity. Another way internal validity is assessed is through inter-rater reliability measures, which helps bolster both the validity and reliability of the study. External validity refers to whether the results of a study generalize to the real world or other situations. A lot of psychological studies take place in a university lab. Therefore, the setting is not very realistic. This creates a big problem regarding external validity. Can we say that what happens in a lab would be the same thing that would happen in the real world? For example, a study on mindfulness involves the researcher randomly assigning different research participants to use one of three mindfulness apps on their phones at home every night for 3 weeks. At the end of three weeks, their level of stress is measured with some high-tech EEG equipment. This study has external validity because the participants used real apps and they were at home when using those apps. The apps and the home setting are realistic, so the study has external validity. See More: Examples of External Validity Concurrent validity is a method of assessing validity that involves comparing a new test with an already existing test, or an already established criterion. For example, a newly developed math test for the SAT will need to be validated before giving it to thousands of students. So, the new version of the test is administered to a sample of college math majors along with the old version of the test. Scores on the two tests are compared by calculating a correlation between the two. The higher the correlation, the stronger the concurrent validity of the new test. Predictive validity refers to whether scores on one test are associated with performance on a given criterion. That is, can a person's score on the test predict their performance on the criterion? For example, an IT company needs to hire dozens of programmers for an upcoming project. But conducting interviews with hundreds of applicants is time-consuming and not very accurate at identifying skilled coders. So, the company develops a test that contains programming problems similar to the demands of the new project. The company assesses predictive validity of the test by having their current programmers take the test and then compare their scores with their yearly performance evaluations. The results indicate that programmers with high marks in their evaluations also did very well on the test. Therefore, the test has predictive validity. Now, when new applicants' take the test, the company can predict how well they will do at the job in the future. People that do well on the predictor variable test will most likely do well at the job. Statistical conclusion validity refers to whether the conclusions drawn by the authors of a study are supported by the statistical procedures. For example, did the study apply the correct statistical analyses, were adequate sampling procedures implemented, did the study use measurement tools that are valid and reliable? If the answers to those questions are all "yes," then the study has statistical conclusion validity. However, if the some or all of the answers are "no," then the conclusions of the study are called into question. Using the wrong statistical analyses or basing the conclusions on very small sample sizes, make the results questionable. If the results are based on faulty procedures, then the conclusions cannot be accepted as valid. Criterion validity is sometimes called predictive validity. It refers to how well scores on one measurement device are associated with scores on a given performance domain (the criterion). For example, how well do SAT scores predict college GPA? Or, to what extent are measures of consumer confidence related to the economy? An example of low criterion validity is how poorly athletic performance at the NFL's combine actually predicts performance on the field on gameday. There are dozens of tests that the athletes go through, but about 99% of them have no association with how well they do in games. However, nutrition and exercise are highly related to longevity (the criterion). Those constructs have criterion validity because hundreds of studies have identified that nutrition and exercise are directly linked to living a longer and healthier life. There are so many types of validity because the measurement precision of abstract concepts is hard to discern. There can also be confusion and disagreement among experts on the definition of constructs and how they should be measured. For these reasons, social scientists have spent considerable time developing a variety of methods to assess the validity of their measurement tools. Sometimes this reveals ways to improve techniques, and sometimes it reveals the fallacy of trying to predict the future based on faulty assessment procedures. References Cook, T.D. and Campbell, D.T. (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin, Boston. Cohen, R. J., & Swerdlik, M. E. (2005). Psychological testing and assessment: An introduction to tests and measurement (6th ed.). New York: McGraw-Hill. Cronbach, L. J. (1970). Essentials of Psychological Testing. New York: Harper & Row. Cronbach, L. J., and Meehl, P. E. (1955) Construct validity in psychological tests. Psychological Bulletin, 52, 281-302. Simms, L. (2007). Classical and Modern Methods of Psychological Scale Construction. Social and Personality Psychology Compass, 2(1), 414 – 433. Validity refers to whether or not a test or an experiment is actually doing what it is intended to do. Validity sits upon a spectrum. For example: Low Validity: Most people now know that the standard IQ test does not actually measure intelligence or predict success in life. High Validity: By contrast, a standard pregnancy test is about 99% accurate, meaning it has very high validity and is therefore a very reliable test. There are many ways to determine validity. Most of them are defined below. Face validity refers to whether a scale "appears" to measure what it is supposed to measure. That is, do the questions seem to be logically related to the construct under study. For example, a personality scale that measures emotional intelligence should have questions about self-awareness and empathy. It should not have questions about math or chemistry. One common way to assess face validity is to ask a panel of experts to examine the scale and rate it's appropriateness as a tool for measuring the construct. If the experts agree that the scale measures what it has been designed to measure, then the scale is said to have face validity. If a scale, or a test, doesn't have face validity, then people taking it won't be serious. Conbach explains it in the following way: "When a patient loses faith in the medicine his doctor prescribes, it loses much of its power to improve his health. He may skip doses, and in the end may decide doctors cannot help him and let treatment lapse all together. For similar reasons, when selecting a test one must consider how worthwhile it will appear to the participant who takes it and other laymen who will see the results" (Cronbach, 1970, p. 182). Content validity refers to whether a test or scale is measuring all of the components of a given construct. For example, if there are five dimensions of emotional intelligence (EQ), then a scale that measures EQ should contain questions regarding each dimension. Similar to face validity, content validity can be assessed by asking subject matter experts (SMEs) to examine the test. If experts agree that the test includes items that assess every domain of the construct, then the test has content validity. For example, the math portion of the SAT contains questions that require skills in many types of math: arithmetic, algebra, geometry, calculus, and many others. Since there are questions that assess each type of math, then the test has content validity. The developer of the test could ask SMEs to rate the test's construct validity. If the SMEs all give the test high ratings, then it has construct validity. Construct validity is the extent to which a measurement tool is truly assessing what it has been designed to assess. There are two main methods of assessing construct validity: convergent and discriminant validity. Convergent validity involves taking two tests that are supposed to measure the same construct and administering them to a sample of participants. The higher the correlation between the two tests, the stronger the construct validity. With divergent validity, two tests that measure completely different constructs are administered to the same sample of participants. Since the tests are measuring different constructs, there should be a very low correlation between the two. Internal validity refers to whether or not the results of an experiment are due to the manipulation of the independent, or treatment, variables. For example, a researcher wants to examine how temperature affects willingness to help, so they have research participants wait in a room. There are different rooms, one has the temperature set at normal, one at moderately warm, and the other at very warm. During the next phase of the study, participants are asked to donate to a local charity before taking part in the rest of the study. The results showed that as the temperature of the room increased, donations decreased. On the surface, it seems as though the study has internal validity: room temperature affected donations. However, even though the experiment involved three different rooms set at different temperatures, each room was a different size. The smallest room was the warmest and the normal temperature room was the largest. Now, we don't know if the donations were affected by room temperature or room size. So, the study has questionable internal validity. Another way internal validity is assessed is through inter-rater reliability measures, which helps bolster both the validity and reliability of the study. External validity refers to whether the results of a study generalize to the real world or other situations. A lot of psychological studies take place in a university lab. Therefore, the setting is not very realistic. This creates a big problem regarding external validity. Can we say that what happens in a lab would be the same thing that would happen in the real world? For example, a study on mindfulness involves the researcher randomly assigning different research participants to use one of three mindfulness apps on their phones at home every night for 3 weeks. At the end of three weeks, their level of stress is measured with some high-tech EEG equipment. This study has external validity because the participants used real apps and they were at home when using those apps. The apps and the home setting are realistic, so the study has external validity. See More: Examples of External Validity Concurrent validity is a method of assessing validity that involves comparing a new test with an already existing test, or an already established criterion. For example, a newly developed math test for the SAT will need to be validated before giving it to thousands of students. So, the new version of the test is administered to a sample of college math majors along with the old version of the test. Scores on the two tests are compared by calculating a correlation between the two. The higher the correlation, the stronger the concurrent validity of the new test. Predictive validity refers to whether scores on one test are associated with performance on a given criterion. That is, can a person's score on the test predict their performance on the criterion? For example, an IT company needs to hire dozens of programmers for an upcoming project. But conducting interviews with hundreds of applicants is time-consuming and not very accurate at identifying skilled coders. So, the company develops a test that contains programming problems similar to the demands of the new project. The company assesses predictive validity of the test by having their current programmers take the test and then compare their scores with their yearly performance evaluations. The results indicate that programmers with high marks in their evaluations also did very well on the test. Therefore, the test has predictive validity. Now, when new applicants' take the test, the company can predict how well they will do at the job in the future. People that do well on the predictor variable test will most likely do well at the job. Statistical conclusion validity refers to whether the conclusions drawn by the authors of a study are supported by the statistical procedures. For example, did the study apply the correct statistical analyses, were adequate sampling procedures implemented, did the study use measurement tools that are valid and reliable? If the answers to those questions are all "yes," then the study has statistical conclusion validity. However, if the some or all of the answers are "no," then the conclusions of the study are called into question. Using the wrong statistical analyses or basing the conclusions on very small sample sizes, make the results questionable. If the results are based on faulty procedures, then the conclusions cannot be accepted as valid. Criterion validity is sometimes called predictive validity. It refers to how well scores on one measurement device are associated with scores on a given performance domain (the criterion). For example, how well do SAT scores predict college GPA? Or, to what extent are measures of consumer confidence related to the economy? An example of low criterion validity is how poorly athletic performance at the NFL's combine actually predicts performance on the field on gameday. There are dozens of tests that the athletes go through, but about 99% of them have no association with how well they do in games. However, nutrition and exercise are highly related to longevity (the criterion). Those constructs have criterion validity because hundreds of studies have identified that nutrition and exercise are directly linked to living a longer and healthier life. There are so many types of validity because the measurement precision of abstract concepts is hard to discern. There can also be confusion and disagreement among experts on the definition of constructs and how they should be measured. For these reasons, social scientists have spent considerable time developing a variety of methods to assess the validity of their measurement tools. Sometimes this reveals ways to improve techniques, and sometimes it reveals the fallacy of trying to predict the future based on faulty assessment procedures. References Cook, T.D. and Campbell, D.T. (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin, Boston. Cohen, R. J., & Swerdlik, M. E. (2005). Psychological testing and assessment: An introduction to tests and measurement (6th ed.). New York: McGraw-Hill. Cronbach, L. J. (1970). Essentials of Psychological Testing. New York: Harper & Row. Cronbach, L. J., and Meehl, P. E. (1955) Construct validity in psychological tests. Psychological Bulletin, 52, 281-302. Simms, L. (2007). Classical and Modern Methods of Psychological Scale Construction. Social and Personality Psychology Compass, 2(1), 414 – 433.