We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Validity in Rehabilitation Research: Description and Classification

Zohreh Salimi and Martin W. Ferguson-Pell

Additional information is available at the end of the chapter

Abstract

There is a considerable body of literature on research validity across different disciplines. With regard to rehabilitation research, however, this body is narrow and warrants further consideration. Classification of research validity has been considered and developed over the past six decades; however, a literature search returned no comprehensive discussion that has gathered all available classifications under an overarching umbrella. The aim of this chapter is to provide an all-inclusive classification of different types of research validity, focusing on rehabilitation research. A basic review of the body of literature available was conducted. Different classifications of validity in the literature were recognized, considered, unified, and are presented in this chapter. Moreover, the main threats to each type of validity and some strategies to minimize them are discussed. A classification of all types of validity in rehabilitation research is presented in this chapter. Furthermore, the matter of priority between these research validities is discussed. It is concluded that while all types of research validities are important to be considered, maximizing all of them in one research project is sometimes controversial. Thus, researchers should make a situation-based trade-off between different aspects of validity in order to optimize the overall validity of their research.

Keywords: research validity, construct validity, classification, threats to validity, rehabilitation research

1. Introduction

Getting concerned with the validity of a research is ensuring its empirical integrity [1]. The level of validity of a study is an indicator of cost-effectiveness and accountability of it [1]. A research study is considered valid when (1) it is able to correctly find the relationship between the variables, (2) it measures what it claims to be measuring, (3) the findings are generalizable,

and also (4) it has adequate statistical power to reject a false null hypothesis. On the other hand, the power of a study is its ability to find the truth: correctly rejecting a false-null hypothesis or supporting a true-null hypothesis. Therefore, a valid study is also a powerful study, because its findings are the true results of that study.

All researchers need to ensure the validity of the tests and instruments they use before conducting research studies, but this is of particular importance for rehabilitation researchers: in a review study, 100 data-based studies were reviewed and by a *post hoc* power calculation, it was shown that there is a high possibility for the occurrence of type II error in rehabilitation research studies [1]. Type II error refers to supporting a false-null hypothesis, which leads to a false-negative conclusion. In the rehabilitation studies reviewed in this study, the medians of power for detecting small-, medium-, and large-effect sizes have been only equal to 0.08, 0.26, and 0.56, respectively. These low-power values clearly show the importance of accounting for power and validity of the study in rehabilitation research, before starting to conduct one. This is because the purpose of all studies is truly discovering the relationship among variables [2], and if a study does not have validity, the results obtained from it will not be trustworthy.

Validity of the research should be considered for all research studies. Research studies concerned with validity have been done for a large variety of different fields of knowledge, including Education [3–12], Psychology [13–16], Marketing [17], Management [18], Employment [19], Nursing [20], Criminal issues [21], Animal studies [22], Sports [23], Nutrition and Food Science [24], Health [25–27], and Rehabilitation Science [1, 2]. Regarding the number of research studies about validity that are available in the literature for different disciplines, Education, Psychology, and Marketing are the fields that are most prominent. However, the body of literature in this context for health studies is narrow, and even narrower for rehabilitation science. This highlights the necessity of addressing this issue for health sciences in general, and rehabilitation sciences in particular. In this chapter, the validity of different types of rehabilitation studies will be discussed. In doing so, the main focus will be on one type of validity that is a more complicated aspect of validation: construct validity. We will discourse some threats to each type of validity along with providing some strategies to prevent them, and hence how to power up the study, as we pass through each validation type.

2. What is validity?

Validity principles are applicable to all studies, whether they are based on questionnaires, observational studies, or other types of assessments [7]. Research validity helps us to know how true the claims and propositions made in a study are. This judgment can be based on the characteristics of a study, such as the research design, adequacy of sample size, the recruitment procedure, instruments and tests used, and the appropriateness of statistical methods used [28].

3. Types of research validity

Research validity can be categorized into two types: internal validity and external validity [29]. Internal validity refers to the amount of credit that can be attributed to the relationship between variables that is true, and external validity refers to how generalizable are the findings. In another approach [28], research validity has been divided into four types: internal validity, statistical conclusion validity, construct validity, and external validity. These four types of validity address four basic questions that practicing researchers face and so it is more practical. Those basic questions are as follows:

(1) Is there any relationship between two variables? (Referring to internal validity)

(2) If so, is it a causal relationship or it might just have happened by chance and can occur without any treatment? (Referring to statistical conclusion validity)

(3) What are the concepts that are involved in this causal relationship? (Referring to construct validity)

(4) How generalizable is this relationship to other settings, tools, persons, and time? (Referring to external validity)

The second classification is actually drawn by dividing each of the first classification components in two [28]: the statistical conclusion validity is differentiated from internal validity to distinguish between the relationship that is affected by covariates and the true relationship between the two variables that are of interest. In other words, internal validity takes care of the validity of the relationship obtained between the two main variables of interest, while the statistical validity makes sure that this relationship is not contaminated (and if so, is correctly taken care of) by other variables that may influence the relationship, but are not of the interest in the study.

Also, construct validity is differentiated from external validity to make a distinction between generalization of the constructs of cause and effect to higher-order constructs and generalization of the findings to the other settings and the population. In other words, the second classification explicitly states what was implicitly covered in the first classification.

The second classification of validity is now widely accepted and is being used by researchers in different fields [1, 2]. We will briefly introduce each of these types of validity in the following sections of this chapter. Furthermore, some threats to each type of validity as well as some strategies to power up that validity will be discussed.

3.1. Statistical conclusion validity

Statistical conclusion validity can be defined as the approximate precision of interpretations concluded about the covariations based on the statistical methods used and the fitness of the research design [2]. In other words, any statistical issue that could influence the results of the

study would be a threat to statistical conclusion validity, including small sample size, lack of statistical power for the study [1], and any random error that could happen by chance, despite appropriate use of statistics, for example, type I and type II errors [30]. These threats to validity happen when the research conditions and statistical process of the study are not rigorous enough [22]. This type of validity is the *most important* type of validity, but in rehabilitation research, it has received little attention [1].

Some of the main threats to statistical conclusion validity are [2, 28] as follows:

- **Low statistical power:** This will reduce the power of the study to reject the null hypothesis. To prevent this threat, researchers should define their experiment characteristics, for example, sample size and eligibility criteria, in order to provide an acceptable statistical power for the research.

- **Violated statistical assumptions**: Each statistical procedure is based on some assumptions, for example, normal distribution of population, homogeneity of variance, and linearity. If these assumptions are violated, those statistical procedures will not be credible anymore. To prevent this threat, the researcher should ensure that the underlying assumptions are met, for example, normality of sample data.

- **Performing multiple statistical tests on a single data set**: This increases the likelihood of type I error and is another threat to statistical validity. This is because the Alpha level of the study as a whole is the sum of the Alpha level of all comparisons made. To prevent this threat, researchers can reduce the number of comparisons by a careful preplanning. Another technique is to determine the Alpha level of each *t*-test in a way that the cumulative Alpha would be desired.

- **Lack of inter- and intra-rater reliability of the study**: Lack of consistency in conditions of implementing experiments is a threat to validity.

- **Lack of reliability of measures:** This would result in both type I and type II errors (e.g., they might show no correlation between two variables, when there is a good correlation, and vice versa). To prevent this threat, researchers must only use reliable measures or tools. In fact, reliability is a necessary condition for validity [14].

3.2. Internal validity

Internal validity deals with whether the treatment used in the research study has an actual effect on the outcome variable [30]. Thus, internal validity is the extent to which we can be confident that there is a certain type of relationship (e.g., causal) between the dependent and the independent variables of the study [2]. If a study has lack of internal validity, then there are some factors in that study, other than the independent variables, that affect the outcome to some extent, but they have not been accounted for [2, 30]. These unaccounted factors are threats to internal validity. There are many threats to internal validity which can be applicable to different types of research. Some of the most common threats are [2, 28, 30] as follows:

- **Maturation:** When experiments get lengthy, the results may be influenced by the participants getting older, wiser, healthier, or stronger. Maturation is considered a threat to validity when it has not been considered in the research design and is not accounted for.

- **History:** When subjects' reaction in the experiment has been influenced by some events that have happened prior to the experiment, for example, when the participant is studied to observe his/her attitude to people with disabilities after some treatments, but in fact his/her past experiences of encountering a person with disability would affect the results.

- **Lack of inter- and intra-rater reliability** will also affect the results and so the causal relationship concluded from those results.

- **Selection**: The researcher should assure that what makes a difference between control and treatment groups is the only factor that is under study. Any other variables that might influence the outcome should not systematically differ between groups. The best protection against this threat is randomly assigning participants to treatment and control groups when possible.

- **Attrition and mortality:** Dropouts that usually happen most for the group that is receiving the harder treatment would influence the outcome, since those who comply with the treatment are generally those who are healthier or are more enthusiastic about the research study they have participated in. Also, mortality can potentially happen for every experiment that needs the attendance of participants for more than one time, but especially it may happen for studies where the participants have serious diseases, such as cancer or heart disease.

- **Sharing information**: If by any means participants have a chance to share their information regarding the experiment, they might influence each other's thoughts and outcomes. This is even more important when the experiment is based on questionnaires and other qualitative methods.

If participants know about other participants in a different group, they may get dissatisfied with the group that they are in and with the treatment type they are receiving and thus they may become disappointed and less motivated to appropriately continue with the treatment (e.g., performing some exercises), which may affect their compliance. Any protection against chances of getting cues about other participants will help to prevent sharing of data. Any strategies that can be taken against hypothesis guessing by participants and also performing a good concealment for the randomization will also contribute to this aim.

3.3. External validity

External validity refers to the population that the study findings can validly be generalized to. In other words, the greater the population the findings can be applied to, both in number and in diversity, the higher the external validity of that study. This, however, depends significantly on the sample used in the study and the population that this sample is actually representing. A simple example of for this is that if you are aiming for generalizing your findings to all wheelchair users but you only recruit male wheelchair users, you have a problem in external validity of your research. Also, the instruments used in the study and other conditions of the

experiments have roles in determining the population the study findings can be generalized to and so in defining the external validity of the study.

This type of validity addresses the generalizability of the findings, which is of particular importance for rehabilitation practitioners, because they need to make inferences from the sample under the study to the treatment provided to a greater population [2]. For making those inferences, we either should perform the same experiment on in different occasions in time, settings, participants, and so on, or one should perform a systematic review with meta-analysis on the body of literature on a particular issue [30]. However, randomly selecting participants in a study (i.e., random sampling) will provide the best protection against threats to external validity of that study and therefore makes the findings generalizable to the population [2]. Some of the main threats to external validity are [2] as follows:

- **Sample characteristics** (e.g., age, gender, race, education, urban versus rural) restrict the population that the findings can be generalized to. Samples should be a good representative of the desired population. Random sampling helps considerably to this aim, but it does not guarantee that this threat would be eliminated.

- **Intervention characteristics** also restrict the findings to the settings with similar features, for example, instruments. Some strategies to prevent this threat include making the use of different examiners (when intra-rater reliability is realized) and using multiple measures that are taken from multiple setting.

- **Context characteristics:** There are some conditions that may influence the way subjects react or respond in the experiment, which inhibit generalization of the findings to the situations with different conditions. For instance, some participants try to provide "correct" responses, which are responses that they believe examiners like to see, but are not representative of their real state. Moreover, sometimes subjects receive multiple treatments at the same time, which may restrict the findings to those people that are on similar treatment regimes.

- **Sensitization:** In research designs where participants receive the same assessment (e.g., questionnaire) pre- and post-test, their knowledge on the way they will be evaluated might affect the outcomes. Since this situation might not be the same as when there is no sensitization on the construct under study, this is a threat to external validity.

3.4. Construct validity

For performing powerful studies and also proper clinical accomplishments, we need to make use of robust measures, considering the fact that "science rests on the adequacy of its measurement" [14]. Using proper and robust measures pertains to construct validity. Construct validity deals with whether an instrument or measurement tool or a test is measuring what it claims to be measuring [11]. In other words, construct validity concerns whether the measurement obtained is really representing the underlying construct [14].

The matter of validity is analogous to a study that has a clear hypothesis; researchers should gather as much evidence as they can to prove the hypothesis about validity of the inference

[11, 13, 14]. Researchers should continue gathering convincing evidence until they feel that they have a large enough set of evidence to prove the construct validity. There is no best way to validate a study, although there are several methods in use [11]. Up to four subclasses of construct validity have been defined: face validity, content validity, criterion-related validity, and construct validity. Some researchers [11, 14, 30, 31] believe that all these subclasses should be grouped and gathered under one overarching umbrella which is the construct validity. This is called the unified view of construct validity. Each of the subclasses of construct validity will be discussed below.

3.4.1. Face validity

This is the first judgment about the validity of an instrument by just looking at the appearance of it. It is only guesswork and provides little evidence for validity of the instrument. Besides, some snags can happen when talking about face validity [14]: the fallibility of verdicts that are based on appearance, different interpretations of appearance between developers and users, and some occasions that the judgment based on the appearance of an instrument is contrary to its contents. Therefore, using only face validity is never sufficient.

3.4.2. Content validity

This type of validity concerns the items or elements of a measurement and the extent to which these elements reflect the area they are supposed to be measuring [31, 32]. The adequacy and fitness of each element of the measurement tool in measuring the targeted construct is discussed under content validity. In other words, the targeted construct guides selecting the content of an assessment tool, and on the other hand the content and elements of the assessment tool selected define the construct that is actually being investigated [32]. Content validity is particularly important in assessing the validity of questionnaires.

Using an instrument that is invalid due to content, results in erroneous conclusions because some aspects of the construct are not represented properly, whether underrepresented or overrepresented. Not accounting for content validity in the study could also result in inaccuracy in finding a significant treatment effect [31].

Content validity is dynamic in nature, because the domain and definition of constructs change by time, and accordingly the elements of an instrument should be changed to be representative of that construct [31]. Content validity of an assessment instrument is dependent on the function of the instrument, population under study, and the situation in which the instrument is used. Therefore, close attention should be taken in order to maintain an acceptable validity for the assessment instrument. Often, a panel of experts is contacted to judge content validity of an assessment or instrument [11, 31].

One example of threats to content validity are is occasions where the definition of one term in researchers' language is different from the commonly accepted definition of it. In this situation, the readers' interpretations of the contents, results, and reports of the study might be

different from the author's and researcher's intent. In these cases, a proper clarification of the constructs is advised [2].

3.4.3. Criterion-related validity

In criterion validity, correlation of the instrument or assessment with a "criterion" is examined [11]. Criterion has to be a "superior" measure that is more accurate than the measure being evaluated; otherwise, the failure in validation might be due to a flaw of the criterion, itself [32]. There are two types of criterion validity [11, 32]: concurrent and predictive, which are introduced briefly, here as follows:

- *Concurrent validity*: This pertains to situations that both the assessment tool that is being tested for its validation and the criterion are measured at the same time. For instance, when blood pressure is measured simultaneously using cuff measurements and intra-arterial pressure measurement tools [32].

- *Predictive validity:* When the assessment tool is tested for its validity by checking how well it can predict a criterion that will happen later. For instance, how well the scores of a test obtained by a sample of people predict their job status in the future [32]. Diagnosis, physiological data, and tests performed in laboratories are examples of instances that predictive validation should be used [32].

3.4.4. Construct validity

In construct validity, we experimentally investigate whether a construct is actually measuring what it claims to be measuring [11]. The concept of construct validity emerged when researchers realized there are many occasions that there is not any "superior" criterion to correlate the instrument under validation study with it [32]. Two types of validity can be distinguished within construct validity [32]:

- *Convergent validity*: Convergent validity is used when in validating a method of measurement, the correlation between that measure and a different method of measurement is assessed, while they are used to measure the same construct [33].

- *Discriminant (divergent) validity*: When we experimentally show that our assessment tool being tested for its validity produces results that are different from data produced by another assessment tool that is measuring another construct and thus should produce different results [33].

Lack of construct validity causes two deficits: [30]

- Contamination: When the scores obtained by the assessment tool represent features that are not part of the construct being studied.

- Deficiency: When there are aspects of the construct being studied that are not included in the assessment tool.

Some threats to construct validity are listed hereafter [2, 11, 13]:

- **Using non-reliable tools**: When reliability is threatened, the construct validity of the device can also be threatened.

- **Narrow stimulus sampling**: When the researcher studies a narrow sample or situation while the construct under study is much broader. Case studies are particularly subject to this threat.

- **Single operations:** When the construct is complex, but the measure inspects just one aspect of it. For example, the researcher measures only the "time spent with friends" as the indication of being happy. To avoid this threat, the researcher should make use of more indicators for the construct.

- **Single subject design:** This design is a threat to construct validity when it is used to implement an intervention or a treatment. This is because the individual's specifications may be responsible for the outcome resulted, not the intervention itself.

- **Experimenter expectancies**: It happens when a researcher has passion and some expectations about the outcome, in a way that it influences his/her interpretation and explanation of the results. This may cause an alternative explanation of the relationship between variables to be drawn, which decreases the construct validity, since it is not declaring the real circumstances of the construct under study.

- If by any means **participants get some clues about the study**, it will affect their performance in the experiment. This is because participant might presume the hypothesis or objective of the study, and act, in the sake of the objective, differently from their usual real behavior. This, in fact, changes the construct that has been assessed, since the study is assessing them when they are "motivated." To prevent or minimize this threat, researchers should attempt to provide fewer cues for participants.

- The same situation is established when participants from one group have **the tendency to compete** with the participants of the other group. This also makes them more motivated. To avoid this threat, researchers should minimize the incidental contacts between subjects.

- **Demoralization**: this happens when some individuals from one group are not satisfied with the treatment they are receiving, compared to the treatment that the other group receives. This makes them less motivated in participating and affects their performance. A solution for this threat is providing another valued treatment for the group that is suspected of being demoralized. Of course, this treatment should be a placebo or proved to have no intervening effect on the study.

- **Mono-method bias**: Happens when the researcher uses only one method of measurement. If this measurement has a poor construct or content validity, the study would be flawed. A method for minimizing this threat is using a number of methods at the same time, for example, questionnaire, self-report, and observation.

- **Mono-operation**: Happens when just one manipulation is used to affect the construct. For instance, to investigate the effect of a special drug as an intervention, participants are

divided to one placebo group and one treatment group. A more solid design will establish multigroup and sets different dosage of the drug for each group.

- **Poor construct definition:** This is when the construct is misdefined (e.g., assessing anxiety instead of depression), or has not been defined properly (e.g., assessing job satisfaction to represent overall happiness). Researchers should get advice from experts in the field before starting the study, to prevent this threat.

Figure 1 depicts a summary of this paper: the overall classification of validities and main issues that threaten different types of validity.
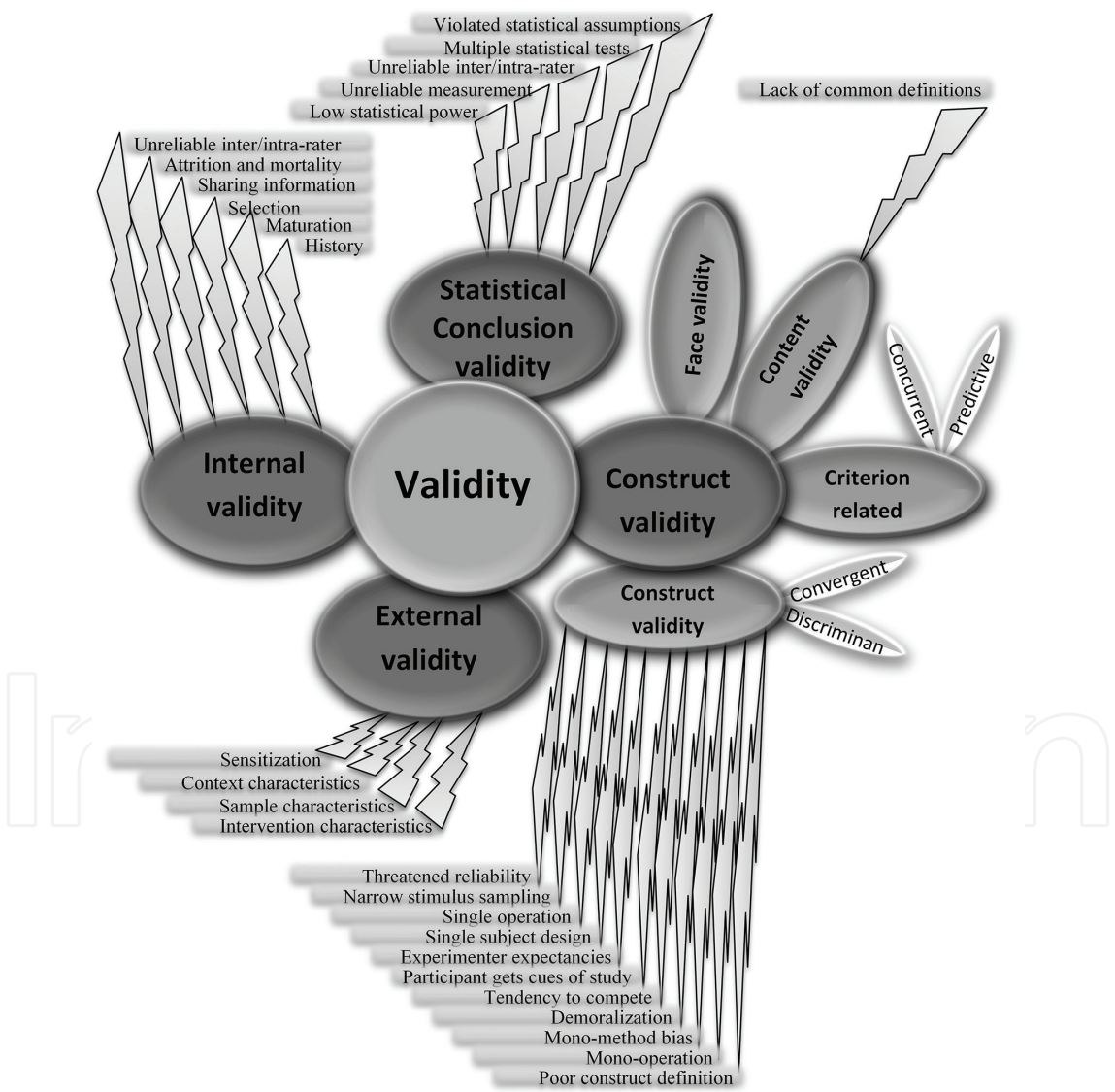

Figure 1. Overall classification of validities and main issues that threaten different types of validity (sparks).

4. Discussion

In this paper, four types of research validity with their subclasses were described. Now, one may ask, among these different types of validity, which one of them is more important and has priority in consideration in validity study? The answer is rather complicated due to different opinions existing in the literature, which will be described here.

Among different types of validity, construct validity is the one that more frequently has been subject to consideration, research, and publication (e.g., see [11, 13, 18]). This is because it appears that for some researchers, construct validity (and its subclasses) are the only concepts of research validity. Although this frequency of consideration relative to the other types of validity could be an indication of relatively greater importance for this type of validity, the authors could not find any explicit declaration of that.

However, as it was pointed out in the "statistical conclusion validity" section, Ottenbacher and Barrett Kathryn [1] believe that this type of validity (statistical conclusion validity) is the most important type of validity, though it has received little attention in rehabilitation research. The importance of this type of validity is that one should make sure that the findings are obtained as a result of a real covariation between variables, rather than chance.

Cook's and Compbell [28] and Mitchell [30], however, do not believe so, since they declare that the internal validity is the most important validity, and hence one should be more concerned about it. This is because this group of researchers believes that as long as a study does not have internal validity, the data achieved from it are not appropriate and trustworthy, and as a result they are not eligible to be generalized to other situations. Mitchell says that most authors are not concerned with external validity and they just put it under "further research" heading [30]. Bellini and Rumrill gather between these two former opinions for studies that investigate an unknown relationship between variables [2].

Some researchers, however, argue that at the end of the day, we need to generalize our data to other situations, and hence if they are valid but unable to be generalized, they are useless. They, therefore, believe that external validity deserves higher priority than what it has received in health research so far [34].

Another point that is worth noting here is the point drawn by Shadish et al. that construct validity is not a necessary condition for external validity, because "we can generalize across entities known to be confounded albeit less usefully than across accurately labeled entities" [35].

Bellini and Rumrill [2] state that in fact, all four research validities are important in turn, and one should try to have all of them in the higher level possible, but there is a logical order for them: first statistical conclusion validity should be established, to show that two variables covary. The second validity to be established is the internal validity, which focuses on the obtained relationship between variables. Then the researcher should be concerned with the construct validity which speaks about the construct that is involved in the relationship.

Eventually, the importance of external validity would arise that is concerned with the generalization of the results to other settings.

To sum up, it should be noted that all validities are of great importance of value for all research. However, one may not be able to maximize all of them at the same time, since maximizing one of them could be dependent on decreasing the other one. For instance, for increasing statistical conclusion validity with a given sample size, one may need to restrict the population under study, in order to decrease the variation between samples and increase the statistical power of the study. This, in turn, obviously leads to a reduction of generalizability of his/her results (lower external validity). Researchers, therefore, need to make trade-offs between different aspects of research validity—considering the conditions they are in, so that they could end up with the optimum validity for their research study.

Acknowledgements

Author details

Zohreh Salimi and Martin W. Ferguson-Pell*

*Address all correspondence to: martin.ferguson-pell@ualberta.ca

Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Alberta, Canada

References

[1] Ottenbacher KJ, Barrett Kathryn A. Statistical conclusion validity of rehabilitation research. Am J Phys Med Rehabil. 1991;**70**(1):138–143.

[2] Bellini JL, Rumrill PD. Perspectives on scientific inquiry: Validity in rehabilitation research. J Vocat Rehabil. 1999;**13**:131–138.

[3] Messick S. The interplay of evidence and consequences in the validation of performance assessments. Educ Res. 1994;**23**(2):13–23. DOI: 10.3102/0013189X023002013

[4] Messick S. Validity and washback in language testing. Lang Test. 1996;**13**(3):241–256. DOI: 10.1177/026553229601300302

[5] Messick S. Meaning and values in test validation: The science and of assessment. Educ Res. 1989;**18**(2):5–11. DOI: 10.3102/0013189X018002005

[6] Messick S. The standard problem meaning and values in measurement and evaluation. Am Psychol. 1975:955–966.

[7] Messick S. Standards of validity and the validity of standards in performance assessments. Am Educ Res Assoc. 1995;**14**(4):5–8.

[8] Messick S. Validity of physiological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol. 1995;**50**(9):741–749.

[9] Shepard LA. The centrality of test use and consequences for test validity. Educ Meas Issues Pract. 1997:5–13.

[10] Jonsson A, Svingby G. The use of scoring rubrics: Reliability, validity and educational consequences. Educ Res Rev. 2007;**2**(2):130–144. DOI: 10.1016/j.edurev.2007.05.002

[11] Brown JD. What is construct validity? Shiken: JALT testing and evaluation SIG newsletter. October 2000:8–12.

[12] Moss PA. Can there be validity without reliability?. Educ Res. 1994;**23**(2):5–12. DOI: 10.3102/0013189X023002005

[13] Shuttleworth M. Construct validity. Explorable. 2005;DOI: 10.4135/9781412950558.n102

[14] Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;**119**(2):166.e7–166.e16. DOI: 10.1016/j.amjmed.2005.10.036

[15] McKay JR, Alterman AI, McLellan AT, Snider EC, O'Brein CP. Effects of random versus nonrandom assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. J Consult Clin Psychol. 1995;**63**(1):70–78.

[16] Roberts BW, Kuncel NR, Shiner R, Caspi A, Goldberg LR. The power of personality: the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspect Psychol Sci. 2007;**2**(4):313–375. DOI: 10.1111/j.1745-6916.2007.00047.x

[17] Brennan L, Voros J, Brady E. Paradigms at play and implications for validity in social marketing research. J Soc Mark. 2011;**1**(2):100–119. DOI: 10.1108/20426761111141869

[18] Arthur W, Woehr DJ, Maldegen R. Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct-related validity paradox. J Manage. 2000;**26**(4):913–835. DOI: 10.1177/014920630002600410

[19] Hunter JE, Schmidt FL, Hunter R. Differential validity of employment tests by race: A comprehensive review and analysis. Psychol Bull. 1979;**86**(4):721–735. DOI: 10.1037//0033-2909.86.4.721

[20] Lynn MR. Determination and quantification of content validity. Nurs Res. 1985;**35**(6): 382–386.

[21] Longshore D, Stein JA. Self-control in a criminal sample: An examination of construct validity. Criminology. 1996;**34**(2):209–228.

[22] Hoffmann A, Skalsky JR. Inferential properties of an individual-based survival model using release-recapture data: sample size, validity and power. J Appl Stat. 1995;**22**(5&6):579–595.

[23] Dunbar CC, Robertson RJ, Baun R, et al. The validity of regulating exercise intensity by ratings of perceived exertion. Med Sci Sports Exerc. 1992;**24**(1):94–99.

[24] Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. J Psychomatic Res. 1985;**29**(I):71–83.

[25] Coste J, Judet O, Barre O, Siaud JR, Cohen de Lara A, Paolaggi JB. Inter- and intraobserver variability in the interpretation of computed tomography of the lumbar spine. J Clin Epidemiol. 1993;**47**(4):375–381.

[26] Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke. 2007;**38**(3):1091–1096. DOI: 10.1161/01.STR.0000258355.23810.c6

[27] Faes MC, Reelick MF, Perry M, Olde Rikkert MGM, Borm GF. Studies with group treatments required special power calculations, allocation methods, and statistical analyses. J Clin Epidemiol. 2012;**65**(2):138–146. DOI: 10.1016/j.jclinepi.2011.05.007

[28] Cook TD, Compbell DT. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Chicago: Rand McNally; 1979.

[29] Campbell DT, Stanley JC. Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin Company; 1963.

[30] Mitchell TR. An evaluation of the validity of correlational research conducted in organizations. Acad Manag Rev. 1985;**10**(2):192–205.

[31] Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. Am Psychol Assoc. 1995;**7**(3):238–247.

[32] Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health Serv Res. 1976;**11**(4):478–507.

[33] Peter JP. Construct validity: a review of basic issues and marketing practices. J Mark Res. 1981;**18**(2):133–145.

[34] Steckler A, McLeroy KR. The importance of external validity [Editorial]. Am J Public Health. 2008;**98**(1):9–10. DOI: 10.2105/AJPH.2007.126847

[35] Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for General Causal Inference. Boston: Houghton Mifflin Company; 2002.