

Using Operational Definitions in Research:

A Best-Practices Approach

Brent D. Slife, Casey D. Wright, and Stephen C. Yanchar

Brigham Young University

## Abstract

The use of operational definitions, though examined philosophically, has not been sufficiently examined from a practical perspective. The practice of operationalization offers obvious benefits to empirical researchers but suffers from a lack of attention to what has been referred to as *translation validity*. Because the relation between an operational definition and its underlying construct can never be measured, the quality of translation validity must be established through conceptual argumentation as well as more traditional means such as converging operations and historical precedent in the literature. More specifically, we suggest that any use of operational definitions should involve best practices related to three conceptual tasks: (a) *clarification*, in which researchers reflect on and clarify their potential operationalizations, (b) *specification*, in which researchers specify and take account of the difference between the construct of interest and what was actually studied via operational definitions, and (c) *justification*, in which researchers assess and defend the translation validity of their particular operationalizations.

Using Operational Definitions in Research:  
A Best-Practices Approach

One of most widely conducted method practices in psychology is one of the least examined—operationalizing. Virtually every psychological method text considers operationalization, or the use of operational definitions, to be a *necessity* for the proper conduct of psychological research. Bordens and Abbot (1999), for example, are straightforward, “... without using operational definitions, questions cannot be answered meaningfully” (p. 30). Similarly, other method texts assert that psychological researchers “must operationalize” (Furlong, Lovelace, and Lovelace, 2000, p. 63; Krathwohl, 2009, p. 141), and that rigorous studies “need” or “require” operationalization (Borg and Gall; 1989, p. 65; Krathwohl, 2009, p. 140) because an operational definition “...gives meaning to a variable...” (Kerlinger and Lee, 2000, p. 43; see also Privatera, 2014, p. 89).

Yet, perhaps surprisingly given its pervasiveness, operationalization has received relatively little critical examination. As we will attempt to show, such examinations have been “sporadic” or “rarely voiced” (Feest, 2005, p. 131; Shean, 2013, p. 74), with most treating operationism as a philosophy of science that is evaluated negatively due to its connection with positivism or post-positivism (e.g., Bickard, 2001; Leahey, 1980; Michell, 2013). Even so, there is no unanimity in such criticism. Feest (2005), for example, has challenged the supposed connection between operationism and positivism.

Our purpose here is more practical than philosophical. We recognize that practical concerns are philosophically situated, but in this article we are interested more in *operationalization* as a practice than in *operationism* as a philosophy. Given the widespread adoption of operational definition (or operationalization) and the scarcity with which it is

examined in practice, we believe that psychological researchers should establish a best-practices approach to the use of operational definitions. As Furlong et al. (2000) noted, "... researchers must be *extraordinarily concerned* with selecting operational definitions and measurement procedures that actually measure what they intend to study.... (emphasis added, p. 64). Our intention, then, is to provide an initial set of recommendations for establishing best practices regarding this "extraordinary concern."

Consequently, we first set the context of this aim by providing some clarifications and a brief history of this research practice. We specifically minimize the connection to philosophies of science because those philosophical issues have already been addressed (Koch, 1992; Leahey, 1980/1981/1983/2001; Michell, 2013), and because the analyses offered are sometimes less than helpful to practicing researchers. Instead, we provide what could be viewed as a kind of commonsense discussion of potential problems with operational definition, clarifying how the practice of conceptualizing operationalizations occurs in the process. Second, we offer an approach to operationalization that may begin a constructive conversation about the limits and uses of this important and relatively unexamined methodological practice. With others (e.g., Bickhard, 2001; Koch, 1992; Leahey, 1980), we agree that there are reasons to question the use of operational definitions in a good deal of psychological research. As we will suggest, however, there are better and worse ways to engage in this practice, provided that researchers' have adequately established a need for operationalization in a given study.

#### Brief History and Clarification

This brief account can only outline some major developments in the use of operational definitions, with other articles describing this history more thoroughly (e.g., Feest, 2005; Koch, 1992; Smith, 1997). Historians of psychology, such as Viney and King (2003), have fairly

routinely credited the physicist Percy Bridgman with “set[ting] forth the principles of operationism” (p. 302) in his classic book (1927) *The Logic of Modern Physics*. However, as Holton (2005) and Walter (1990) have described, Bridgman was also one of the first to doubt the usefulness of operational definition to psychology. These histories also clarify that although physics might have influenced psychology in the beginning, the most important recent psychological developments of operationism have occurred within psychology itself.<sup>1</sup>

S. S. Stevens and Edward Tolman were two of the first psychologists to popularize the use of operational definitions (Feest, 2005, p. 131), with the first significant critical debate following their contributions in the 1940s (Bergmann and Spence, 1941; Israel, 1945; Israel and Goldstein, 1944; Pennington and Finan, 1940; Waters and Pennington, 1938; Weber, 1940). Two more recent periods of debate occurred in the early 1980s and early 2000s, with Leahey (1980) and Kendler (1981) as important figures in the first, and Grace (2001) and Bickhard (2001) as principals of the second. These debates, however, were of a “philosophical nature” and, as we suggested above, primarily concerned possible connections to philosophies of science (Feest, 2005, p. 132).

Feest (2005), as one of the more recent contributors to this literature, has questioned the “conceptual and historical assumptions” (p. 132) of these debates, arguing in particular that the concerns of the original champions of operational definition in psychology—Stevens and Tolman—were less about philosophies and more about practicalities. Feest, in fact, questions whether this more recent literature about the philosophies of operationism have had “much relevance to operationism *as practiced by psychologists*” (emphasis in original, p. 132). She even describes how Stevens’ and Tolman’s own references to philosophies of science, positivism in particular, were provided *after* the fact of their original need and use of operational

definitions. In other words, their own claims about specific philosophies may have had less to do with their practices than even they realized.

It is with actual research practice that the present article is concerned. Given the possibility that some psychologists have misrepresented their *own* practices through potentially erroneous references to philosophies of science, we believe it is important that the discipline focus on practical method procedures, particularly as set forth in the method texts used to train researchers. Again, we do not question that such practices are to some degree embedded in a scientific worldview, but the examination of operational definition has focused almost exclusively on philosophical concerns. Also, there are good reasons to question *in practice* whether many psychological phenomena are measurable in any defensible sense, or whether operationalized versions of phenomena are the most viable articulations of what researchers wish to investigate (see analyses by Michell, 1999, 2003, Stam, 2006). As we will suggest, whether or not operationalization is strategically beneficial in a given situation must be carefully considered, and it is likely that often it is not. Nonetheless, when the use of operational definitions can be justified, a guiding sense of how they are best formulated and implemented would offer significant benefits in terms of rationale and rigor. Thus, it is high time, we believe, to formulate a *best-practices* approach to operationalization, especially as the practice is described in research methods texts.

These texts present the practice of operational definitions in two prominent ways—the first involving the purpose or intention of the researcher for deploying this procedure, and the second involving its actual application as a scientific operation. Both conceptions are considered parts of the practice of using operational definitions. Regarding the purpose of operational definitions, there is virtual unanimity among the method texts of psychology as to how

researchers are trained to think about operationalization— its purpose is to provide clear and specific scientific measurement (McBride, 2013; Morling, 2015; Nestor and Schutt, 2015; Passer, 2014; Privitera, 2014). Hoyle, Harris, and Judd (2002), for example, describe how operationalizations should “specif[y] precisely how to measure a variable in such a concrete and specific manner that anyone else could repeat the steps and obtain the same measurements” (p. 76). Consider also how Morling (2015) expresses this purpose: “to turn a concept of interest into a measured or manipulated variable” (p. 57).” As Durlak and DuPre note, “Science cannot study what it cannot measure accurately and cannot measure what it does not define” (2008, p. 342). Consequently, investigators must “translate abstract concepts that cannot be directly observed into tangible, measurable variables” (Passer, 2014, p. 116).

But how specifically are researchers trained to apply the purpose of translating “abstract concepts” into “tangible” measurements? We can only provide here a fraction of what our survey of method texts yielded, but there is almost complete agreement about what one concretely does to establish an operationalization (McBride, 2013; Passer, 2014; Privitera, 2014). The essential practice or procedure is to translate the construct of interest that is not measurable or observable into something related to it that *is* measurable or observable. Consider, for example, how the *APA Dictionary of Statistics and Research Methods* (2014) treats this issue: one operationally defines when one translates the concept into “terms of the operations (procedures, actions, or processes) by which it could be observed and measured” (p. 245). Or consider Privitera’s (2014) claim that the purpose of operationalizing is to conceptualize “some observable event in terms of the specific process or manner by which it was observed or measured” p. 565). And Kerlinger and Lee (2000, p. 43-44) sum up this practice by stating that there “... can be no scientific research without observations ...” because an observable operation “... assigns meaning to a

variable...”.

At this point, many psychologists may even assume that these two understandings of operationalization—its purpose and its practice—are synonymous, i.e., that clear, scientific specification and measureable operations are one and the same. However, as many qualitative approaches to method show, one may have the purpose of clarifying and specifying without necessarily making the unmeasured quantitatively measureable. In a prominent qualitative research text, for example, Marshall and Rossman (1999) recommend that research questions be “general enough to permit exploration” but also “focused enough to delimit the study” (p. 38), because, as other qualitative researchers suggested, the question “becomes progressively narrowed... during the research process” (Strauss and Corbin, 1990, p. 37). As a case in point, consider how Adams (2015) began her qualitative study with a “larger analysis” of women’s overall chronic illness but then “facilitate[s] clarity” by “narrowing” to a more specific analysis of “posttraumatic growth,” which is not itself measurable in the conventional, quantitative sense (p. 15). Our point in making this distinction is that purpose (specification) and practice (making topics measureable) are not identical; typical quantitative practices are only one way to seek clarity or specification.

This distinction is useful for other reasons as well. For example, when scholars claim that “all empirical psychologists have to operationalize their concepts” (e.g., Feest, 2005, p. 145; M. Freeman, personal communication, August 9, 2014), we would need some elaboration about what this claim means. It could mean that all broadly empirical psychologists—both qualitative and quantitative investigators—are concerned about clarifying and specifying their subject matter for investigation. We believe that most, if not all qualitative and quantitative researchers would endorse this claim. However, it could also mean that all broadly empirical psychologists

attempt to turn the unmeasured into the quantitatively measurable, which we believe many qualitative researchers would *not* endorse (Marshall and Rossman, 1999; Stake, 2010; Strauss and Corbin, 1990).

### The Central Issue

With this brief history and clarification as background, we can begin a discussion of what some would call the “mainstream” practices of operationalization, the practices of many quantitative and perhaps a few qualitative psychologists, especially if they were trained through method texts to translate unobserved constructs into observable, measurable operations.

#### *Translation Validity*

A best-practices approach would focus on Krathwohl’s (2009) notion of “translation validity,” which is “the closeness with which the study’s intended meaning of constructs matches their operationalization” (p. 405). As Hoyle et al. (2002) note, operational definitions are “never completely adequate” because they “rarely seem sufficient to capture the rich and complex ideas contained in a construct” (p. 76). Furlong et al. (2000) put the validity issue in this manner: “common sense should tell you that if we fail to measure the right things or if we fail to measure them the right way, we will be unable to answer the right question” (p. 63). In this sense, pivotal to a best practices approach is an evaluation of a study’s translation validity, a relatively overlooked subset of construct validity.

But why might we expect that translation validity varies, at least to some degree, across studies? We believe it is obvious that some operational definitions can be good and some bad, especially when the researchers themselves formulate such definitions for particular studies (Morling, 2015, p. 123). Some might assume that the peer review process weeds out bad operationalizations. They may believe that reviewers will dislike certain operationalizations

because of their poor quality and thus reject those studies, allowing only those with the best operational definitions to be published. This is likely the case, to some degree, but there are two problems with this process. First, reviewers have no established criteria for assessing translation validity. Although some evaluation of operationalizations surely still occurs, guidelines for such evaluations are not available. Second, examples of questionable translation validity abound in published research. Here we offer an example of a published study that has been criticized for what could amount to translation validity issues. We do not single this study out because it is unique. Indeed, as we will describe, we would contend not only that this investigation is a relatively good one but also that many studies are potentially subject to the same types of criticisms.

The study in question is titled “The Neural Correlates of Hate,” authored by reputed neuroscientists at University College, London (Zeki and Romaya, 2008), and published in *Plos One*, a highly respectable and impactful journal. Participants brought in photos of people they “hated” as well as photos of people they felt “neutral” toward. By comparing the participants’ brain activity while viewing each set of photos, the researchers claimed to have identified the neurological correlates of hatred. Subsequent media portrayed this identification as “Hate Circuit Found” (Tibbetts and Brealey, 2008; Robson, 2008), and the lead author, Semir Zeki, contended that such results would likely be used in court to evaluate the state of mind of murder suspects (Robson, 2008).

In a brief critique, Satel and Lilienfeld (2013) looked at the findings of the Zeki and Romaya (2008) study differently. Although Satel and Lilienfeld seemed to admit that their data do reveal the brain activity associated with the hated photos, the problem is that “the illuminated areas on the scan are activated by other emotions, not just hate. There is no newly discovered

collection of brain regions that are wired together in such a way that they comprise the identifiable neural counterpart of hatred” (p. 32). Satel and Lilienfeld do not attribute this problem explicitly to the translation validity of the operationalization, but they are clearly referring to Krathwohl’s (2009) definition of this term: “the closeness with which the study’s intended meaning of constructs matches their operationalization” (p. 405). They sum up the validity issue this way:

It’s all too easy for the nonexpert to lose sight of the fact that fMRI and other brain-imaging techniques do not literally read thoughts or feelings. By obtaining levels of brain-oxygen levels, they show which regions of the brain are more active when a person is thinking or feeling, or, say, reading or calculating. But it is a rather daring leap to go from these patterns to drawing confident inferences about how people feel about political candidates or paying taxes, or what they experience in the throes of love (p. 33).

As we will detail in the next section, Satel and Lilienfeld are describing—however implicitly—one of several potential validity problems with operational definitions: many such definitions can miss what the investigators mean or intend for them to be, allowing for alternative interpretations of the data. In a similar sense, hugs and kisses would often be considered a good operational definition for love, yet any data relevant to this operationalization could also pertain to unwanted advances or even a Mafioso death. This is not to say that studies of patterns of hugs and kisses or patterns of brain activity are not valuable. Our point is that there are clear translation validity questions regarding the closeness of the operationalization to the meaning of the intended construct that need to be addressed.

### *The Main Problems*

Why, then, are some operationalizations good and some bad? If we can get clear about

the main problems, perhaps we can begin to suggest possible best practices in the use of operational definitions. In this article, we can only begin a discussion of potential problems, but we would like to formalize three that many researchers are likely to have already sensed informally, even though these problems are rarely discussed in method texts. We also use as an illustration Tolman's classical operational definition of rat hunger: time since feeding (Feest, 2005; Tolman, 1932).

*Problem 1: Operational definitions are not identical to their constructs.* Problem 1 is merely the claim that changing the focus of study (translating it) from something abstract and possibly unobservable to something that is measurable and observable, is to alter, however slightly, what is *actually* studied. As obvious as this problem may seem, few researchers address it in any substantive manner. Zeki and Romaya (2008), for example, are not atypical among psychological researchers of any stripe when they represent the variable they are studying as "hate" (note their title above) rather than the oxygenation of certain brain regions, which is what they are *actually* studying. In fact, it is apparent throughout their paper that they treated the construct of their study as identical to the operationalization they actually studied—a common method practice in psychology—when the assumption of this identity relation is obviously not a minor issue, as Satel & Lilienfeld (2013) note.

Consider Tolman's operationalization of hunger in this regard. As close a match as time-since-feeding would appear to be to rat hunger, it seems quite possible that some rats could be hungrier than others when the dependent variable is measured (e.g., slightly different rat constitutions, some rats more active than others). Even if rat constitution and rat activity were controlled or equated in some manner, differences in hunger among the individual rats, perhaps even systematic differences, cannot be ruled out because their *actual* hunger cannot be measured.

We could even say that their actual hunger cannot be measured *in principle*, because the inability to measure actual hunger empirically is the very reason the operationalization is needed. In this sense, if problem 1 is correct, what Tolman was studying cannot be considered identical to hunger. Discussing his study as if it is all about rat “hunger” could be misleading, even if only slightly—hence the issue of translation validity.

This validity issue may grow in importance as we move to humans and less controllable circumstances. Investigations of human love, for example, could lead to many potential operationalizations. We mentioned hugs and kisses above, but we believe the issue is the same with any operationalization. Hugs and kisses are simply not identical to love. Hugs and kisses can occur without love, and love can occur without hugs and kisses. This was Satel and Lilienfeld’s point about the neural correlates of hate: the brain activity revealed on the scan could refer to some other emotion. In fact, even Zeki and Romaya (2008) concede that the same region of the brain has been associated with love emotions (p. 6).

Is it also possible for the hate emotion to occur without the brain activity specified by Zeki and Romaya (2008)? There is surely little doubt that the brain is somehow involved in most forms of hate, but could some forms involve other parts of the brain, or perhaps even other portions of the nervous system? The answer to this question would depend on how the researcher conceptualized the type of hate as well as the complexity of factors associated with each type. The hate of someone who murders someone else, Zeki’s claim, could be quite different from those who “hate” exercise but do it every day. Moreover, the fMRI profiles about hate were originally reliant on the subjective report of the “hate” of the first participants, allowing inevitable variability between participants. In other words, there are many variables in translation validity, from all the possible meanings of the construct to all the possible

formulations and measurements of the proposed operationalizations (e.g., see Fabianssen, Moulds, Grisham, and Schira, 2012 on anger).

It is also important to note that Problem 1 does not stem from what some have termed the “essentialist critique” (e.g., Stanovich, 2013, pp. 37-52). From Stanovich’s framing, an essentialist would insist on some ultimate (or essential) meaning of the construct being operationalized. In this perspective, the constructs of hunger, love, or hate would have a single ultimate or fundamentally true meaning, which each operational definition would need to approximate, if not represent in its entirety, to be valid. However, Problem 1 does not require some *ultimately true* meaning of the construct in question. As far as Problem 1 is concerned, researchers can investigate whatever they might construe as love, hate, or hunger, without requiring a particularly true or correct meaning. This problem involves the closeness between *whatever* the researcher’s construct of interest might be and its operational definition. How well does the latter represent the former? For example, a researcher’s construct for love may have little to do with romantic love per se and more to do with a grandmother’s love for her grandchildren. However, translation validity remains an issue because whatever the operationalization is, it is not identical to grandmotherly love.

We recognize that some investigators may claim the prerogative to simply *identify* the construct with the operationalization, such as identifying hunger with time-since-feeding or hate with the specified brain activity. However, this move is merely translation validity by fiat. With this prerogative, *anything* could be declared as hunger or hate, depending on the whim of the investigator, which is surely not a scientific approach to the issue.

Similarly, some researchers may want to study only the operationalization, that is, only the behaviors they can observe, detecting possibly important patterns in their data with no

pretense as to its identity with a construct. Tolman, in this sense, could only have been interested in studying time-since-feeding, and Zeki and Romaya may only have been interested in studying a particular pattern of brain activities. We do not wish to comment on the significance of such studies, as we have already suggested. Our only concern is the translation validity issue—the closeness of match between the construct and its operationalization. Again, we are not developing an essentialist position here, where the construct can only mean a certain thing, a definitive definition. However, a construct does not imply just anything. The researcher typically has some meaning in mind, especially when most constructs can have multiple meanings, implying a more specific understanding. In this sense, focusing on the operationalization (measurement) only is not a problem as long as researchers do not assume some relation with a construct as they interpret data and report results. Failing to exercise such caution could lead at least to potential misrepresentation, where findings about hugs and kisses are erroneously assumed to reveal something important about the nature of love itself, whatever type of love the researcher has in mind. As we will also see, such assumptions are even more problematic with problem 2.

*Recommendation.* The best practice in light of problem 1 is relatively simple—research reports *cannot* assume that the construct itself is being studied in any sort of measurable or straightforward manner through its operationalization. Instead, authors of such reports need to specify, as precisely as possible, what was *actually* studied, while discussing explicitly how what was studied might have been different, however slightly, from the original construct of interest. Consider also how the disparity between construct and operational definition may affect the implications and conclusions of the study. If the operational definition differs substantially from the intended construct, then one should not draw implications and recommendations about the

intended construct per se from the results, as tempting as it may be to do so. However, as we review the other problems of operationalization (below), we will see how certain research justifications can be used to provide credence for the particular choice of operationalization, and thus some degree of confidence that results based on the operational definition are applicable to the construct of interest to some degree. It is possible that the need to formulate such justifications may also lead to better operationalizations.

*Problem 2. Constructs that require operationalization are not measurable in principle, so their relationship to the operationalization is not directly measurable.* Returning to our Tolman example, problem 2 means that the relation between rat hunger and the actual time-since-feeding is not itself empirically measurable (i.e., not observed or counted). Because Tolman could not actually observe the rat's hunger—however he might have conceived of this construct—he will never measure directly the degree to which his conception of this unobserved state is related, if at all, to the empirical findings of his study of the operationalization. The reason is again straightforward: the relationship of unmeasurable phenomena (construct) to measurable phenomena (operationalization) cannot be empirically measured or observed because part of that relationship is not itself measured or observed.

The point is the same for both the hate/brain activity relationship and the love/hugs relationship. Again, researchers may claim the prerogatives discussed above: either eliminating the construct and studying the brain activity alone or identifying the hate *with* the brain activity. For the same reasons described in the previous section, however, neither prerogative eliminates problem 2. The first is not an issue of translation validity because nothing is being translated; the second is translation validity by fiat, something that is typically viewed as outside the usual practices of scientists.

The point of problem 2, then, is that no direct measurement is possible to support the translation validity of a *particular* operationalization. We emphasize “particular” operationalization here, because we will discuss the case of multiple operationalizations, either within or across studies, in problem 3. The issue here is that the unmeasurable nature of the particular relationship between the construct and its operationalization obviates direct empirical evidence. Indeed, if it were possible, there would be no need for the operationalization in the first place.

*Recommendation.* The clear implication of problem 2 is that there is no best practice of direct measurement in providing translation validity for a particular operationalization’s relationship to its construct. However, this implication does not mean that a persuasive conceptual case cannot be made. In many instances, the very logic of the study goes to its conceptual persuasiveness. For example, the logic of the Zeki and Romaya (2008) study is strengthened if participants followed the experimenters’ instructions and actually brought hated and neutral photos. Moreover, these researchers attempted to discern each participant’s feelings about the hated person through a paper-and-pencil test of their own devising, called the Passionate Hate Scale (PHS).

The implication here is that the case for the translation validity of a particular operationalization has to be made conceptually. While it is true that Zeki and Romaya (2008) report empirical covariance between the PHS and the brain activity, the meaningfulness of these correlations relies greatly on their conceptual relationship. If, for example, the meaning of the PHS ratings were in no way conceptually related to the hate meaning of the participants, a spurious correlation could still occur. These correlations might be akin to infamous spurious correlations such as the divorce rate in Maine and its per capita consumption of margarine ( $r =$

.99; <http://www.tylervigen.com/spurious-correlations>). Obviously, higher correlations do not necessarily mean better correlations, unless the correlations themselves make rational or conceptual sense. For this reason, the logic of the methods (e.g., internal and external validity) comes into play for providing important levels of operational credentials and thus translation validity.

*Problem 3. Because we cannot directly measure the construct's relation to its operationalization, we cannot directly measure the relationship among different operationalizations to the same construct.* This may be the most challenging problem to understand, not to mention overcome. In fact, multiple operationalizations have routinely been cited as the best way to establish translation validity, whether through “literature validity” (across previous studies) or convergent operationalizations (within a study) (e.g., Grace, 2001). The use of multiple operationalizations appears to follow from the assumption that statistical correlations among different operationalizations indicate translation validity—that is, they should indicate a meaningful conceptual relation among those operationalizations and the construct of interest. Covariance, in this sense, is taken to provide evidence of an underlying identity relationship between the construct and its operationalizations. In this sense, the most persuasive evidence for translation validity is the formulation of a set of empirically demonstrable, internally consistent, and convergent operations.

Assuming an underlying identity between multiple operationalizations and a construct can be the eventual basis of a justification for the use of those operationalizations. However, such a justification does not resolve the central issue stated in Problem 2—namely, that we cannot directly know empirically the relationship between a construct and its operationalizations. The lack of directly measurable knowledge with one operationalization in Problem 2 is only

multiplied with several, unmeasurable relationships between the construct and operationalizations in Problem 3. In other words, if researchers cannot empirically observe and measure a relationship between an abstract construct and a given operationalization—and thus do not know if an operationalization is a valid translation of that construct—they also cannot empirically observe and measure a relationship between an abstract construct and multiple operationalizations.

Indeed, the multi-operations approach could merely compound a previously invalid inference. A statistical correlation among operationalizations may or may not point to an underlying identity among them—one needs to make a conceptual case for the translation validity of such a (co)relation (see problem 2). In this sense, justification for the validity of any such translation—that is, the translation of construct to operationalization, or the translation of operationalization to operationalization—depends critically on conceptual or rational argumentation regarding why a given operationalization provides the best way to investigate the construct in question. Statistical correlation can be helpful, but it alone will not establish translation validity. Empirically speaking, identity relations can only be inferred on the basis of empirical covariance *and* a logical or conceptually convincing explanation of this covariance.

To return to the example of Tolman's research, he could have operationalized hunger as a particular reduction of bodyweight or quantity of food provided, in addition to time-since-feeding. However, statistical correlations among these operationalizations do not logically imply that these operationalizations are connected, nor does it validly lead to the conclusion that an underlying identity (hunger) actually is reflected among the operationalizations and the construct. What if, for example, some sort of gastric illness had unknowingly afflicted some of the rats but not others? In this case, the afflicted rats could be identical to the non-afflicted rats

in the quantity of food provided and time since feeding, but their reduction in bodyweight, though ostensibly equivalent to the non-afflicted rats, was, in fact, due to the gastric illness (and not to their hunger). Moreover, this illness could also differentially affect the afflicted rats such that they were less (or more) hungry than the non-afflicted rats. The upshot is that the multiplicity of operationalizations does not ensure the translation validity of any or all of the operationalizations. Indeed, assuming that such covariance among operationalizations *automatically* ensures greater confidence in translation validity could be greatly misleading.

This is not to say that multiple operationalizations cannot be helpful in justifying or building a case for a study's translation validity. Hypothetically speaking, Tolman could certainly have seen converging operations as evidence of an underlying identity among operationalizations and his hunger construct. Surely, at least, a reduction in body weight, along with a decrease in food provided, could rule out some alternative explanations of the lone operationalization, time-since-feeding, and thus strengthens the persuasiveness of the study to actually involve rat hunger. However, given the alternative interpretations possible for any operationalization, this justification for translation validity would necessarily involve *both* the covariance *and* the conceptual argumentation needed to do the "ruling out." In this sense, all operationalizations, including multiple operationalizations, need to be discussed and defended, because the covariance alone only points to a measured relation; it cannot point unequivocally to a measured identity (construct). Needless to say, these sorts of issues only become more complex as we move from rat hunger to human hate and love, all the more requiring the researcher's conceptual explanation of translation validity.

*Recommendation.* As in the case of problem two, establishing translation validity for multiple operationalizations cannot depend solely on empirical procedures, because no identity

relation between a construct and its operationalizations can be directly measured. In this sense, translation validity is not solely an empirical matter. Best practices in using operational definitions, then, must be pursued in conjunction with other means. Most pertinently, it includes the formulation of an accompanying argument, including the ruling out of potential alternative explanations as to why a given set of operationalizations provides the most defensible way to collect empirical data related to the construct in question. The validity theorist, Michael Kane (2013), advocates a similar idea, demonstrating that validation in general (test score interpretations, uses, etc.) is a matter of providing a coherent argument. In short, all operationalizations require some kind of defense; their translation validity cannot be merely presumed, regardless of the covariance established.

Moreover, we would caution that construct validity is not synonymous with translation validity. Construct validity is thought to be achieved either through what is essentially multiple operational definitions (DeVellis, 2017), which can seriously compound translation validity difficulties (problem 3), or through solely empirical (or statistical) arguments, which frequently lack the necessary rational or theoretical justifications just described. In this sense, translation validity is important to foreground in itself when designing a study

### Discussion

As virtually all mainstream methods texts note, operationalization is an important aspect of psychological research. However, with a few rare exceptions (e.g., Krathwohl, 2009), method texts fail to indicate anything about the validity of individual operationalizations—how it is achieved or even that such validity is needed. Yet, the relationship between the construct that is intended to be studied and the operationalization that is actually studied is critical to scientific

knowledge, at least from this mainstream perspective.

How is this translation validity obtained? The answer is unfortunately complicated by the unmeasurable relation between the construct, which is unmeasurable in principle, and the measurable operationalization. Indeed, not understanding the unmeasurable status of this relation is likely part of the reason that operationalization has not received more attention in research texts and training. The measurability of the operationalization has been taken for granted, especially since measurability is its main purpose. What has been frequently overlooked is its unmeasured relation to the construct that prompted a given study. How then can translation validity be established when the researcher cannot rely solely on empirical data? The short answer is that researchers must build translation validity into the logic of the method design and explicitly address this logic in the report of the study. We will lay out the longer answer in the next section, but let us first consider the hate study example.

As we mentioned earlier, the Zeki and Romaya's (2008) study has been criticized for whether these researchers actually observed neural correlates of hate (Satel and Lilienfeld, 2013). To their credit, however, Zeki and Romaya seem to have anticipated such challenges to the translation validity of their operationalization of hate—the participants viewing hated photographs—so they built into their design *another* operationalization of hate, a questionnaire that assessed the participants' hate feelings about the person in the photo. The study's design did not require this questionnaire, because the sought after neural correlates depended methodologically on the participants viewing the hated photographs. Still, the researchers must have realized intuitively (because they do not report it explicitly) the importance of translation validity for their study: were the study participants actually experiencing hatred when the researchers observed their neural correlates?

In this sense, the questionnaire was an attempt to bolster the translation validity of the hated photographs, because this validity was pivotal to the significance and understanding of the investigation's findings. We do not doubt that some intuitive sense of the need for translation validity occurs among psychological researchers who use operational definitions. If such validity is as important as we claim, why wouldn't it be sensed? Our point is that we should not leave this validity to the chance intuition of researchers; this particular validity should be foregrounded and conceptualized carefully and explicitly.

It bears noting, for example, that the translation validity of both operationalizations in the hate study—viewing the hated photo and taking the questionnaire—is easily attacked if this validity is not addressed explicitly. First, both operationalizations ultimately stem from self-reports, with all the well-known validity issues of this genre of measures. Second, the authors described their own theory of hate that guided their construction of the questionnaire. We applaud the authors for this explication of the questionnaire's underlying theory, but they make no effort to connect their formal theory of hate with the informal theories of the participants whose hate this questionnaire attempted to assess. What if, for example, the quality or meaning of their formal theory of hate differed from what participants experienced when considering the photographs? We do not claim that this necessarily was the case, nor do we claim that the researchers cannot address these issue in some way. Our point simply is that most psychological researchers, including Zeki and Romaya (2008), appear not to be trained to address translation issues, even when these researchers apparently sense the need for such validity.

As Satel and Lilienfeld suggest, we could also be concerned about misrepresentation issues in this study. For instance, the first author reportedly told the press that such brain scans could “assess whether a murder suspect felt a strong hatred toward the victim” (Satel and

Lilienfeld, 2013, p. 32). Satel and Lilienfeld correctly cry foul by noting that the “illuminated areas on the scan are activated by many other emotions, not just hate” (p. 32). Nevertheless, this criticism is significantly less problematic when this study’s operationalizations have translation validity. In other words, if Zeki and Romaya had convincingly addressed the translation validity of their operationalizations, the participant’s hate would have been viewed as more likely producing the neural correlates. Granted, a finer grained analysis of these correlates might still be necessary to separate the emotions associated with these neural regions, but no such fine-grained analysis is possible without close attention to what is actually being studied—translation validity. A persuasive discussion about the translation validity of what was actually studied could go a long way toward indicating that this particular region of the brain is a prime candidate for this particular form of hate.

At this point, we hope it is apparent that the issue of multiple operationalizations does not exempt the researcher from addressing translation validity. The correlation or covariance of such operationalizations, which Zeki and Romaya (2008) demonstrated, does not by itself establish translation validity. Method texts often trumpet the old adage, “correlation does not mean causation,” but they often neglect an equally important statement, “correlation does not mean identity.” In this case, the correlation of two or more operationalizations, along with the researcher’s assertion that this correlation indicates the same construct, is insufficient in itself to establish that these operationalizations relate to the same unobserved construct.

If, for example, participants in the hate study interpreted their instructions to mean more of a mild dislike for the persons in the photos, and the questionnaire assessed the kind of hate that could cause a murder, the two operationalizations could covary without being identical. Indeed, mild dislike and vehement detestation might even involve different neural correlates.

Consider also the two operationalizations of hugs and kisses for the construct of love. These could be highly correlated and yet the first (hugs) could be related to “church friends,” while the second (kisses) is viewed as “romantic” in nature. Science cannot merely take the researcher’s word that two or more correlated operationalizations are the same ultimate identity; this proposed identity needs to be discussed and defended.

For this reason, translation validity depends not only on statistical correlation but also on conceptual plausibility. As previously discussed, conceptually *implausible* correlations abound that approach coefficients of 1. In the example mentioned above, no one reasonably believes that the consumption of Maine margarine is meaningfully related to the divorce rate, at least not until some plausible theoretical connection is proposed. Similarly, operationalizations that covary to some degree cannot be understood to have relevance to the same construct without some conceptual plausibility. Therefore, we hope to begin a new conversation about the explicit establishment of translation validity. We say “begin” here because we fully recognize that our comments can only initiate a longer conversation about how such validity can and should be attained. We thus suggest, cautiously, the following compilation of our recommendations (above) as suggestions for research training and practice.

#### Specific Recommendations for Training and Practice

We have grouped our recommendations, tentatively, under the headings of clarification, specification, and justification. Because of space considerations, these recommendations cannot be a step-by-step, “how to” for researchers. However, we believe that method teachers and text authors will be able to derive important suggestions from this brief description.

*Clarification*

A proper approach to translation validity requires researchers to reflect upon and clarify these operationalization issues in the design phase of the investigation, rather than some later point in the investigation. Indeed, we would contend that translation validity is a relatively overlooked aspect in establishing internal validity, and should be considered accordingly in the formulation of the study's design. How is the validity of the operationalization going to be addressed in the design? There are typically many options for operationalizing constructs. Researchers need to clarify for themselves at the outset why they select some operationalizations over others, and how they intend to justify the particular option they elect (see "Justification" below). Perhaps even the limitations of particular genres of operationalizations (e.g., self-reports) or operationism as a philosophy could be beneficially considered.

We also recognize the tendency for many researchers to adopt some operationalization (or set of operationalizations) from previous published studies—sometimes known as "literature validity." However, the mere assertion that the operationalization is validly connected to the construct of interest, even when published, does not make it so. Unless previous studies have engaged explicitly in discussions that address the design issues of translation validity, literature validity is insufficient. Further, researchers will need to address important context differences between their study and previous studies. Our review of operationalizations across the research literatures suggests that old operationalizations are frequently used without justification in dramatically different studies (e.g., different participants, different interventions).

*Specification*

All reports of an investigation need to specify, as precisely as possible, what was *actually* studied, while discussing explicitly how what was studied might have strayed, however slightly,

from what was *intended* to be studied. Again, we find that some investigators recognize intuitively the importance of this specification. Rarely, however, is enough information provided for the reader to evaluate the construct validity issues at play.

These specifications should first include some description of the researcher's intended meaning for the construct (e.g., the investigator's particular meaning of hate or love or hunger). Given that the meanings of constructs can vary greatly—for example, from hating a certain food to hating a race of people—researchers cannot assume their readers will simply know what types of meaning the researchers are attempting to study. Second, the authors of the report should specify how closely the actual operationalization, the actual observations measured, approximates the intended meaning of the construct of interest. Because operationalizations are not identical to the constructs (Problem 1), some description of obvious similarities and differences between the construct and operationalization is needed.

### *Justification*

At this point, with considerations of translation validity incorporated into the method design (clarification), and obvious differences between constructs and operationalizations described (specification), it is important to justify and defend explicitly the translation validity of an operationalization. We recognize that empirical researchers may not be accustomed to such justifications, but in a real sense the entire logic of scientific investigation is a justification of sorts (Kleiner, 1993; Slife and Williams, 1995). Because the validity of operationalizations is not just a matter of empirical relation, authors of the research report should not avoid actively arguing for the conceptual or rational plausibility of the correlated observations. As discussed above, multiple operationalizations can only be correlated (or covaried), rather than experimentally manipulated, so they are just as vulnerable to accusations of spurious relationship

as any correlation.

Addressing such accusations requires a threefold approach. First, the logic of the research design, given the proper attention to “clarification” above, should address such accusations. For example, the researchers of the hate study, likely sensing the importance of translation validity, added another operationalization of hate to bolster the self-reported hate of the photograph. Second, the quantitative covariance of the operationalizations is vital. Although we have emphasized the insufficiency of such empirical relationships (given that important relations in operationalization are unmeasured), we do not mean to imply that they are unimportant. Obviously, an extremely low covariance of the operationalizations would severely harm the case for translation validity. In this sense, empirical or quantitative relations are *necessary* to translation validity, but they are not *sufficient* in themselves. As mentioned, Zeki and Romaya (2008) demonstrated the covariance of their operationalizations.

What these researchers did not attempt is the third portion of our threefold approach to justification. The authors of any such report need to engage in an active explanation of the plausibility of the operationalization or multiple operationalizations. Why is this operationalization justified on conceptual grounds? Why is the correlation among multiple operationalizations not spurious in nature? One important approach to providing this justification is to review briefly major rival operationalizations and explain the justification for not selecting them, especially in light of the perceived, more plausible operationalizations chosen. For this reason, it is vital when multiple operationalizations are used that their relationship not merely be assumed or asserted.

### Conclusion

Operationism, as a philosophical consideration, has been surrounded by controversy for years, but the use of operational definitions in actual research practice has received little, if any, serious attention. We acknowledge that operationism is suspect on philosophical grounds, and that research based on operational definitions often produces distorted versions of psychological phenomena, rendering results distinct from, and often irrelevant to, what the investigator may have intended to study with the initial construct being operationalized (Koch, 1992; Leahey, 1980). For this reason, operational definitions should be used with caution. One caution concerns whether or not operational definitions should be used at all in a given study. Operational definitions are not only different from the construct chosen for study; this difference (between the unobserved construct and the observed operationalization) is not empirically knowable because this relation is itself unobserved. Consequently, we suspect that for investigations of many phenomena, methods not based on operationalization would be more capable of offering meaningful findings. When operational definitions are used, it seems reasonable that a case should be made regarding their necessity as a way of gaining access to the phenomena being studied. A second caution, which we have primarily addressed here, concerns ways that researchers who use operational definitions could more carefully and rigorously engage in this practice.

As we have argued, the practice of operationalization suffers from a lack of attention to what has been referred to as *translation validity* (Krathwohl, 2009). Because the relation between an operational definition and its underlying construct can never be observed or measured, the quality of translation validity must be established through conceptual argumentation as well as more traditional means of converging operations and historical precedent in the literature. More specifically, we suggest that any use of operational definitions should involve three conceptual

tasks: (a) *clarification*, in which researchers reflect on and clarify their potential operationalizations, (b) *specification*, in which researchers specify and take account of the difference between the construct of interest and what was actually studied via operational definitions, and (c) *justification*, in which researchers assess and defend the translation validity of their particular operationalizations. Our intention is to raise awareness of the problems that can result from an unexamined use of operational definitions and begin a conversation regarding best practices for researchers whose studies depend on them.

## References

- Adams, H. L. (2015). Insights into processes of posttraumatic growth through narrative analysis of chronic illness stories. *Qualitative Psychology*, 2(2), 111-129.
- Bergmann, G., and Spence, K. (1941). Operationism and theory in psychology. *Psychological Review*, 48, 1-14.
- Bickard, M. H. (2001). The tragedy of operationalism. *Theory and Psychology*, 11(1), 35-44.
- Bordens, K. S., and Abbott, B. B. (1999). *Research design and methods: A process approach*. (4<sup>th</sup> ed.). Mountain View, California: Mayfield Publishing Company.
- Borg, W. R., and Gall, M. D. (1989). *Educational research: An introduction*. (5<sup>th</sup> ed.). White Plains, New York: Longman.
- Bridgman, P. (1927). *The logic of modern physics*. New York, New York: Macmillan.
- DeVellis, R. F. (2017). *Scale development: Theory and applications*. (4<sup>th</sup> ed.). Thousand Oaks, California: Sage Publications.
- Durlak, J. A., and DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350. doi: 10.1007/s10464-008-9165-0.
- Fabiansson, E. C., Denson, T. F., Moulds, M. L., Grisham, J. R., and Schira, M. M. (2012). Don't look back in anger: Neural correlates of reappraisal, analytical rumination, and angry rumination during recall of an anger-inducing autobiographical memory. *NeuroImage*, 59, 2974-2981.
- Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences*, 41(2), 131-149.

- Furlong, N. E., Lovelace, E. A., and Lovelace, K. L. (2000). *Research methods and statistics: An integrated approach*. San Diego: Harcourt College Publishers.
- Grace, R. (2001). On the failure of operationism. *Theory and Psychology*, 11(1), 5-33.
- Holton, G. (2005). "B. F. Skinner, P. W. Bridgman, and the 'Lost Years'," in *Victory and Vexation in Science* (pp. 65-80). Cambridge, Massachusetts: Harvard University Press.
- Hoyle, R. H., Harris, M. J., and Judd, C. M. (2002). *Research methods in social relations* (7<sup>th</sup> ed.). New York: Wadsworth.
- Israel, H. (1945). Two difficulties in operational thinking. *Psychological Review*, 50, 273–291.
- Israel, H., and Goldstein, B. (1944). Operationism in psychology. *Psychological Review*, 51, 177–188.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kendler, H. H. (1981). The reality of operationism: A rejoinder. *Journal of Mind and Behavior*, 2(3), 331-341.
- Kerlinger, F. N., and Lee, H. B. (2000). *Foundations of behavioral research* (4<sup>th</sup> ed.). Orlando, Florida: Harcourt, Inc.
- Kleiner, S. A., 1993, *The Logic of Discovery: A Theory of the Rationality of Scientific Research*, Dordrecht: Kluwer.
- Koch, S. (1992). Psychology's Bridgman vs Bridgman's Bridgman. *Theory and Psychology*, 2(3), 261-290.
- Krathwohl, D. R. (2009). *Methods of educational and social science research: The logic of methods* (3<sup>rd</sup> ed.). Long Grove, Illinois: Waveland Press Inc.
- Leahey, T. H. (1980). The myth of operationism. *Journal of Mind and Behavior*, 1(2), 127-143.

- Leahey, T. H. (1981). Operationism still isn't real: A temporary reply to Kendler. *Journal of Mind and Behavior*, 2(3), 343-348.
- Leahey, T. H. (1983). Operationism and ideology: Reply to Kendler. *Journal of Mind and Behavior*, 4(1), 81-89.
- Leahey, T. H. (2001). Back to Bridgman?! *Theory and Psychology*, 11(1), 53-58.
- Marshall, C., and Rossman, G. B. (1999). *Designing qualitative research* (3<sup>rd</sup> ed.). Thousand Oaks, California: Sage.
- McBride, D. M. (2013). *The process of research in psychology*. Thousand Oaks, California: SAGE Publications, Inc.
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. New York, New York: Cambridge University Press.
- Michell, J. (2003). The quantitative imperative: Positivism, naïve realism, and the place of qualitative methods in psychology. *Theory and Psychology*, 13, 5-31.
- Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology*, 31(1), 13-21.
- Morling, B. (2015). *Research Methods in Psychology: Evaluating a World of Information*. New York, New York: W.W. Norton and Company, Inc.
- Nestor, P.G., and Schutt, R.K. (2015). *Research methods in psychology: Investigating human behavior*. Thousand Oaks, California: SAGE Publications, Inc.
- Passer, M. W. (2014). *Research methods: Concepts and connections*. New York, New York: Worth Publishers.
- Pennington, L., and Finan, J. (1940). Operational usage in psychology. *Psychological Review*, 47, 254-266.

- Privitera, G. J. (2014). *Research methods for the behavioral sciences*. Thousand Oaks, California: SAGE Publications, Inc.
- Robson, D. (2008). 'Hate circuit' discovered in brain. *New Scientist*. Retrieved on August 24, 2015 at <https://www.newscientist.com/article/dn15060-hate-circuit-discovered-in-brain/>.
- Satel, S., and Lilienfeld, S. O. (2013). Losing our minds in the age of brain science. *Skeptical Inquirer*, 37(6), 30-35.
- Shean, G. D. (2013). Controversies in psychotherapy research: Epistemic differences in assumptions about human psychology. *American Journal of Psychotherapy*, 67(1), 73-87.
- Slife, B. D., and Williams, R. N. (1995). *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. Thousand Oaks, California: Sage Publications.
- Smith, R. (1997). *The Norton history of the human sciences*. New York and London: W. W. Norton and Company.
- "Spurious correlations". (2015). Retrieved on September 17, 2015 from <http://www.tylervigen.com/spurious-correlations>.
- Stake, R. E. (2010). *Qualitative research: Studying how things work*. New York, New York: The Guilford Press.
- Stam, H. J. (2006). Pythagoreanism, meaning, and the appeal to number. *New Ideas in Psychology*, 24(3), 240-251.
- Stanovich, K. E. (2013). *How to think straight about psychology* (10<sup>th</sup> ed.). Boston, Massachusetts: Pearson, Inc.
- Strauss, A., and Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, California: Sage.
- Tibbetts, G., and Brealey, S. (2008). 'Hate circuit' found in brain. *The Telegraph*. Retrieved on

August 24, 2015 at

<http://www.telegraph.co.uk/news/newstoppers/howaboutthat/3274018/Hate-circuit-found-in-brain.html>.

Tolman, E. (1932). *Purposive behavior in animals and men*. New York: Century.

Viney, W., and King, D. B. (2003). *A history of psychology: Ideas and content* (3<sup>rd</sup> ed.). New York: Allyn and Bacon.

Walter, M. (1990). *Science and cultural crisis: An intellectual biography of Percy Williams Bridgman (1882-1961)*. Stanford: Stanford University Press.

Waters, R., and Pennington, A. (1938). Operationism in psychology. *Psychological Review*, 45, 414–423.

Weber, C. (1940). Valid and invalid conceptions of operationism in psychology. *Psychological Review*, 49, 54–68.

Zeki, S., and Romaya, J. P. (2008). The neural correlates of hate. *Plos One*, 3(10): e3556.  
doi: 10.1371/journal.pone.0003556.

Zedeck, S. (2014). *APA Dictionary of statistics and research methods*. Washington, DC: American Psychological Association.

---

<sup>1</sup> Indeed, unlike physics, where operationism was “never very influential,” operational definition “gained a fair amount of popularity within psychology and the social sciences” ((Feest, 2005, p. 131; cf. Smith, 1997, p. 668).