International Journal of Aviation Research

Volume 14 | Issue 01 Peer-Reviewed Article

Construct Validity

David Trafimow
New Mexico State University

Validity of measurement is crucial for conducting useful aviation research, and many have emphasized the importance of a particular type of validity, construct validity. However, despite this emphasis, few researchers understand that which construct validity entails, and believe that merely establishing a nice factor structure is sufficient. It is not, as is exemplified by Cronbach and Meehl's (1955) classic insistence on establishing a nomological network that intersects with observations at a variety of points. Thus, the first goal is to explain construct validity in simple language that also clarifies the seldom addressed topic of what a hypothetical construct is in the first place. The second goal, however, is to explore construct validity criticisms and their implications. This section features a potentially competing type of validity termed auxiliary validity, that can be argued more useful than construct validity. However, if researchers are to continue to emphasize construct validity, they should understand what it entails.

Recommended Citation:
Trafimow, D (2022). Construct validity. *International Journal of Aviation Research, 14*(01), 96-110.

Construct validity is an important concept in aviation research, but many aviation researchers seem not to understand it. To address the misunderstanding, the present goals are as follows. The first goal is to explain construct validity in a way that is easy to understand and that will stick. An important part of that is explaining what a hypothetical construct is, but also explaining what construct validity is not. Secondly, however, even with the meaning of construct validity clarified, there remains the issue of how much researchers should care about it. It is quite possible to understand construct validity and believe that it matters, but it also is quite possible to understand construct validity and believe that it does not matter very much. The extent to which construct validity matters depends not only on its own strengths and weaknesses, but also the strengths and weaknesses of alternative validity conceptions. As will be discussed, a relatively new kind of validity, auxiliary validity, might be more useful than construct validity. Thinking in terms of auxiliary validity points to a vulnerability in construct validity that deserves emphasis.

What is Construct Validity?

Hypothetical Constructs

To understand construct validity properly, it is first necessary to understand the notion of a hypothetical construct, which is an explanatory variable that cannot be observed directly. It is easy to provide examples: attitude, behavioral intention, anxiety, ability, intelligence, and so on. Moreover, theories typically causally connect hypothetical constructs. For example, the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein, 1980; Fishbein & Ajzen, 1975; Fishbein & Ajzen, 2010), which has been used in aviation research (e.g., Fussell & Truong, 2020), asserts that attitude is one cause of behavioral intention, neither of which can be observed directly and both of which are hypothetical constructs (hereafter, hypothetical constructs will be shortened to construct or constructs).

Constructs are crucial in the so-called hard sciences too. For example, consider Newton's famous equation: *force = mass* x *acceleration*. *Mass* is a nonobservational term that Newton never defined, and for good reason. Had Newton attempted a definition, he would have had to use more words which, in turn, he would have had to define, thereby leading to an infinite series (Lederman, 1993; Trafimow & Rice, 2009). Instead, Newton used the term as a primitive, not to be confused with *weight*, which is observational. The difference becomes immediately obvious upon considering that a particular object weighed on earth or moon would have the same mass but different weights.

That constructs are not observable is not all. Another characteristic that constructs have is they contain surplus meaning that goes above and beyond the sum of the different ways of manipulating or measuring them. There are many ways to measure *attitude*, but the term has meaning that goes beyond any single measuring technique or even the sum of all measuring techniques thus far invented.

That constructs cannot be observed and have surplus meaning are not disconnected characteristics. The necessity that constructs have surplus meaning forces them to be impossible to observed directly; if constructs could be observed directly, they could be defined by the observation, and there would be no surplus meaning. A way around this is to insist that only observational terms be used in science—an attempt made approximately a century ago by the positivists—but positivism failed for a variety of reasons that extend beyond present scope. Today, philosophers recognize widely that constructs cannot be observed and have surplus meaning that cannot be captured fully by how they are manipulated and measured.

Measures of Constructs

If an aviation researcher wishes to measure relevant constructs, that researcher likely would wish to have valid measures. Textbooks often define a measure as valid if it measures what it is supposed to measure. However, this is not very useful as it begs the question: How can one know if a measure measures what it is supposed to measure? When Cronbach and Meehl (1955) proposed construct validity, the idea was to answer the question. It was already well known that other sorts of validity, such as criterion-based validity, were subject to important limitations. In the case of criterion-based validity, the idea is that the measure of a construct should correlate with the measure of another construct where the two constructs are related. The obvious counterargument is that the correlation could be spurious and so criterion-based validity provides a very weak case for concluding that the measures really measure what they are supposed to measure.

Cronbach and Meehl (1955) came up with the clever idea that it is possible to support the theory and the validity of the measures of the constructs simultaneously. A poorly understood construct validity assumption is that one is doing two jobs, not one job. The goal is to support the theory and that the measures are valid, with both accomplishments proceeding simultaneously. The essence of construct validity can be stated in a single phrase: **construct validity is the matching of theoretical and empirical relations**. If a researcher can show that measures of constructs that are theoretically connected correlate with each other, but that measures of constructs that are theoretically unconnected do not correlate with each other, such demonstrations aid in supporting both theory and validity of measures, i.e., establishing construct validity. It is possible to characterize this process more precisely.

To avoid confusion, I will use capital letters (*A*, *B*, …, *Z*) to refer to constructs and lower-case letters with subscripts (a_1, a_2, a_3, b_1, b_2, b_3, …z_1, z_2, z_3) to refer to measures of constructs. For example, *A* could refer to depression and a_1, a_2, a_3 to various measures of depression, such as the Beck depression inventory (Beck et al., 1961). Suppose that a theory states that *A* causes *B*. If it were possible to observe *A* and *B* directly, and we have already seen it is not, we would expect that the causal effect of *A* on *B* would force the two constructs to correlate, absent complicating factors interfering with the correlation. However, although there is no way to observe *A* and *B* directly, we can observe them indirectly via recourse to measure a_1 for construct *A* and measure b_1 for construct *B*. The

straightforward prediction is that a_1 and b_1 should correlate. This prediction would simultaneously support (1) the theory that *A* causes *B* (or at least that *A* and *B* are related) and (2) the measures are valid. Of course, this one finding would be preliminary and insufficient on its own to make a strong case for either the theory or the validity of the measures.

It is possible to go further. Suppose that we have five measures of *A* and five measures of *B*, and that all the measures of *A* correlate with all the measures of *B*. Clearly, the case for construct validity would improve; there is yet more reason to believe the theory and there is more reason to believe the measures are valid. To go yet further, we might expect the five measures of *A* to correlate extremely highly with each other because they are all measuring the same construct, the five measures of *B* to correlate extremely highly for the same reason, and for the measures of *A* to correlate somewhat highly with the measures of *B* due to their measuring theoretically correlated constructs.

Nor are we done. Consider another construct *X* that is not mentioned by the theory or, better yet, is explicitly not related to *A* or *B*. In that case, although measures of *X* should correlate highly with each other, there is no reason to expect them to correlate highly with measures of *A* or measures of *B*. In this case, a failure to obtain these correlations would simultaneously support the theory and the validity of the measures.

And it is possible to continue to move further. Consider that according to the theory, *A* causes *B*. An implication is that if *A* is experimentally manipulated, there should be an effect on *B*. But how do we know that the manipulation of *A* is valid or that the measure of *B* is valid? It is important to keep in mind that although *A* cannot be observed, a manipulation of *A* can be observed. For example, a researcher could manipulate attitudes with a pro-attitude essay or anti-attitude essay; although attitudes are not directly observable, it is easy to observe the essays themselves. The question is whether the essays validly manipulate attitude. Suppose that a researcher performs a manipulation that is alleged to influence *A* and obtains the predicted effect on a measure of *B*. This empirical success would simultaneously support that *A* causes *B* and that the manipulation of *A* is valid and the measure of *B* is valid. The empirical success is not definitive because the manipulation might have unintentionally manipulated something other than *A* that is causally related to *B*. Nevertheless, the researcher is on stronger ground with the empirical success than without it, though the extent of that increased strength may be qualified by the face validity of the manipulation, the availability of plausible alternative explanations, and other considerations. Of course, the case could be strengthened by measuring constructs other than *A* that researchers propose as alternative explanations and showing that measures of these other constructs are not influenced by the manipulation.

It is possible to go yet further, but I'll trust that the point has been made that there are many ways to attempt to match theoretical and empirical relations. Over several studies, possibly conducted by different researchers, it is possible to establish a network of theoretical relations that intersects at various points with observations. Such a network is sometimes called a *nomological network*. At this point it should be obvious that

construct validity is not dichotomous. Construct validity is not there or not there, but rather supported by a more intricate nomological network or a less intricate nomological network buttressed by contact with more or fewer corresponding observations. An analogy would be a spider's web, where adding strands increases the strength of the whole web. When reading a paper where a researcher asserts that construct validity has been established, it is good to remain skeptical and ask: "Precisely how intricate is the nomological network ostensibly supporting construct validity?" and "At how many points is the nomological network buttressed by corresponding empirical relations?"

What Construct Validity Is Not

Researchers often use factor analysis to develop their measures. The typical assertion, with citations omitted to protect the guilty, is that if a set of items load strongly on one factor, and very weakly or not at all on another factor, this demonstrates construct validity. That is, the items that load on each factor compose construct valid measures of respective factors.

The claim is poor on multiple levels. In the first place, it is possible to ask whether the alleged latent constructs are connected by a theory. If not, construct validity is immediately out of the question because there is no way to relate empirical relations to theoretical relations when there are no theoretical relations. A theory is a prerequisite for construct validity. And asserting that the factor model is the theory is insufficient; to say that the theory implies the model and that the model implies the theory is blatantly vicious.

In addition, remember that a factor is simply a mathematically generated entity based on the correlations or covariances among items. The hope is that there is a latent construct that causes the responses to the various items that load on a factor, but there is no guarantee that this is so. Of course, it is possible, while constructing a structural equation model, to assert that latent constructs cause responses on the various items that load on respective factors, but assertion falls well short of proof or demonstration. This seems a good place to repeat the truism in introductory classes in psychology, marketing, management, economics, medicine, and others: correlation need not imply causation.

Moreover, even if a researcher establishes a beautiful factor structure, and even if we ignore the foregoing comments, that factor structure is no substitute for showing many points of contact between a nomological network and corresponding empirical relations. This requires, at minimum, an impressive pattern of a variety of types of empirical victories as described earlier. Establishing a beautiful factor structure is, at best, a beginning, but it is not equivalent to that which is required any more than finding a dollar is equivalent to achieving millionaire status.

This is not to say that factor analysis is irrelevant to construct validity. For example, sometimes a theory will specify the factors that ought to come about, in which case establishing the theorized factor structure provides a beginning to establishing construct validity. In addition, an elegant factor structure is usually to be preferred over a

messy one with items loading moderately on different factors. But establishing the factor structure would be insufficient unless followed up by serious attempts to match empirical with theoretical relations. Such attempts might include showing that (a) the factors correlate strongly with other measures of the same constructs, (b) the factors do not correlate strongly with measures of outside constructs or of constructs that might easily be mistaken for the theorized constructs, (c) measures of theoretically related constructs correlate but not as highly as measures of the same constructs, (d) performing experimental manipulations to support that alleged causes of one or more factors play the theorized causal role, (e) performing experimental manipulations to support that the constructs indexed by the factors causally determine other constructs in the theory as indexed by their factors, and (f) others. Factor analysis is a beginning, not an ending, and it is neither necessary nor sufficient for establishing construct validity.

The upshot is that anytime a researcher claims to have established construct validity, or that the cited researchers have established construct validity, the correct reaction is to put your hands in your pockets to determine if your wallet is still there. Despite routine claims of construct validity, it is extremely rare when there is a sufficiently intricate nomological network, buttressed by numerous points of contact with corresponding empirical relations, to support those claims. The Beck Depression Inventory (Beck et al., 1961) provides an example of a rare level of construct validity because it is (a) based on a theory that features negative cognitive schemata, (b) includes a nomological network linking negative cognitive schemata to other constructs, and (c) has resulted in empirical relations that match theoretical relations. There are numerous points of contact between the nomological network and observation documented in hundreds of articles. But even here, there is no such thing as perfection. For one thing, the initial inventory has undergone considerable revision. In addition, even currently, there are criticisms about too much reliance on self-report. Establishing construct validity is extremely difficult—a much more difficult task than researchers typically understand it to be—and there is rarely a strong case for it in published research.

How Much Should You Care About Construct Validity?

I will answer the question right now: I do not know how much you should care about construct validity. The caring issue is complex, and I will attempt to present some, but not all, of the complications.

Of What is the Construct Constructed?

Slaney and Garcia (2015) pointed out an increasing tendency for researchers to be lax about specifying that which constructs are supposed to represent. It is one thing to have a well-articulated set of phenomena and propose a construct to represent that set. It is quite another thing to propose a construct that is based upon little more than hope or intuition. In the present environment, where researchers are increasingly under pressure to include everything that might matter in their studies, there is an unsurprising proliferation of constructs poorly backed in the sense of relating to well-articulated sets of phenomena.

However, it is not clear whether this is really a criticism of the notion of construct validity or if it is rather a criticism of how researchers proliferate constructs with insufficient justification. Although the criticism should not be ignored, it perhaps is not fatal for construct validity as a philosophy of validity. That researchers misuse construct validity need not invalidate the idea of construct validity.

The Circularity Criticism

It is possible to criticize construct validity on the grounds of circularity. One assumes the theory and the measures, finds that empirical relations are consistent with assumed ones, and then concludes that construct validity has been established. Thus, the antecedents (assuming the theory and measures) support the consequents (empirical relations) and the consequents support the antecedents. In a word, there is circularity.

However, it is not clear that the circularity criticism is fatal. It is often true, in basic science, that the research is motivated by a theory, and in turn the findings support the theory. There arguably is a circularity, but it is not clearly vicious, as it is possible for the results to fail to support the theory or to support the validity of the measures.

For example, suppose again the theory that *A* causes *B*. Suppose a researcher attempts a manipulation of *A* and gets an effect on one measure of *B* (b_1) but not another measure of *B* (b_2). There are many possible explanations. It could be that the theory is simply wrong and the effect on b_1 is because the manipulation influenced b_1 without having anything to do with *A*. The manipulation may have influenced some unknown construct that is correlated with b_1, thereby creating an effect on b_1. The theory might be true, with b_1 a valid measure of *B* and b_2 an invalid measure of *B*, hence the effect on b_1 but not on b_2. There are other possibilities too, but these are sufficient to render salient that it is not inevitable that the theory is supported or that the validity of the measures or manipulations is supported. Therefore, the issue of circularity, though potentially relevant, need not be fatal.

The Criticism of Positive Social Outcomes

Messick (1989, 1998) felt that an important defining characteristic of validity concerns the degree to which using the measures of interest produces positive social outcomes. Thus, according to Messick, construct validity might be missing a crucial component because it fails to consider the production of positive social outcomes.

However, admitting the desirability of positive social outcomes is not necessarily fatal for construct validity. One potential defense might invoke the distinction between applied and basic research. Although applied research might be reasonably expected to produce positive social outcomes, basic research is concerned with the production of knowledge, and there is no expectation of an immediate, or even somewhat long-term, bettering of the human condition. An advocate for construct validity could argue that if the process of establishing construct validity produces knowledge gain, in the form of

support for the theory and for the validity of the measures, that is desirable from a basic research perspective even without clarity about the ultimate betterment of the human condition.

Of course, this is only one potential defense, and other defenses need not rest on a distinction between applied and basic research. An alternative defense might be that there is no way to know whether a theory or measure of a construct within that theory will lead to positive social outcomes until someone makes the attempt. Even if that has not happened yet, or even if attempts had been made but failed, a success could happen later, and so it is premature to pronounce a measure invalid just because it has not yet been applied to achieve positive social outcomes. Moreover, there are many examples in the history of science where knowledge that originally was not applied to produce positive social outcomes eventually produced them.

The foregoing counters do not exhaust the possible defenses. But they are sufficient to demonstrate that a lack of demonstration of positive social outcomes is not fatal for construct validity.

Auxiliary Validity

Thus far, we have avoided the issue of how one traverses the distance between nonobservational constructs in theories and observational manipulations and measures. Considering this issue in detail provides a stronger case against construct validity than in the foregoing subsections. A disadvantage of considering this issue is that it requires a long subsection for proper consideration.

It is convenient to commence with the prediction that Edmond Halley [*1656-1742*] made about the year of reappearance of the comet that now bears his name. He used Newton's theory, but he also made additional assumptions about the present position of the comet, the presence or absence of gravitationally relevant bodies, and so on. These latter assumptions are not contained in Newton's theory; they are auxiliary to the theory and are termed *auxiliary assumptions*. The punchline is that auxiliary assumptions are necessary to traverse the distance between nonobservational theoretical terms, such as mass or attitude, and observational empirical terms such as weight or a particular attitude manipulation or measure.

The necessity to have auxiliary assumptions is a problem for the philosophy of falsification. Suppose that a theory predicts a finding, but the finding does not occur, so there is an empirical defeat. Typically, we would like to say that the empirical defeat falsifies the theory (e.g., Popper, 1959), but this is an oversimplification (e.g., Lakatos, 1979; 1978). Rather, we could conclude that the theory is false or that one or more of the auxiliary assumptions was at fault. Similarly, when there is an empirical victory, it is possible to credit the theory, but it also is possible to credit the auxiliary assumptions (Trafimow, 2017). Hence, it is unlikely that a single empirical defeat or empirical victory is definitive.

Once we admit that auxiliary assumptions are crucial for theory testing, other matters follow. Consider that we test theories by performing experimental manipulations and assessing effects on measures of constructs, or by assessing relationships between measures of constructs. Either way, the validity of the manipulations and measures depends on the quality of the auxiliary assumptions (Trafimow, 2012). And there is no escaping that fact.

But if the validity of the manipulations and measures depends on the quality of the auxiliary assumptions, then that implies a different way to look at validity. It becomes immediately self-evident that if the auxiliary assumptions are true, the manipulation or measure is valid, and if the auxiliary assumptions are not true, the manipulation or measure is invalid. I previously coined the term *auxiliary validity* to express the dependence of manipulations and measures on auxiliary assumptions (Trafimow, 2012). Arguably, it is better to think of the auxiliary assumptions as varying on a dimension from high-quality to low-quality to have a continuous concept as opposed to a dichotomous one. However, this issue is not of particular importance for present purposes.

What is important for present purposes is that an admission that auxiliary assumptions are crucial for validity poses an important, and perhaps fatal, problem for construct validity. To see this, consider some examples. One example concerns a theory that predicts that under stated conditions, there ought to be radioactivity present. A researcher sets up the conditions, measures the radioactivity with a Geiger counter, but fails to find the hoped-for radioactivity. The empirical defeat could be because the theory is wrong, because the researcher made wrong auxiliary assumptions and unintentionally failed to set up the stated conditions, or because the auxiliary assumption of a working Geiger counter is wrong. Let us suppose certain knowledge that the theory is wrong and that all auxiliary assumptions are true. Is there construct validity and is there auxiliary validity? With respect to the latter, the answer is in the affirmative: we supposed certain knowledge that the auxiliary assumptions are true and auxiliary validity follows inevitably. About construct validity, there is a failure for empirical relations to correspond with theoretical relations; according to the theory (and auxiliary assumptions), the Geiger counter should have registered the requisite level of radioactivity and it did not. The fact that the study is valid with respect to auxiliary validity, but not with respect to construct validity, suggests that construct validity is problematic. There arguably is no reason to care about construct validity given that we have auxiliary validity.

For a second example, consider that attitudes are supposed to cause behavioral intentions. Suppose a researcher measures attitudes and behavioral intentions and obtains a strong positive correlation, thereby supporting both the theory and the validity of the measures. The researcher has made a nice beginning at establishing a nomological network pertaining to the theory of reasoned action, with points of contact with observation, in the direction of construct validity. One of the auxiliary assumptions, of course, is that the attitude measure really measures attitudes. But Fishbein (1980) famously made a strong case that some (or most) ostensible measures of attitudes

measure variables other than attitudes. Let us suppose that in the present study, attitudes were measured in one of the ways that Fishbein criticized and let us further suppose that Fishbein got the criticism right. In that case, there is an obvious auxiliary validity failure; the auxiliary assumption that the attitude measure really measures attitude is wrong. However, moving to construct validity, recall that the empirical relations correspond nicely with the theoretical ones; there should be a strong correlation between attitude and behavioral intention measures, and there is one. Thus, we again have a mismatch between auxiliary validity and construct validity, thereby again illustrating what might be a fatal flaw for construct validity.

The two examples imply two important questions, bullet-listed below.

- Can one have validity if the theory is wrong?
- Can one have invalidity if the theory is right?

Let us consider both questions under the umbrellas of construct validity and auxiliary validity. Commencing with construct validity, let us remember that there is no commitment to the theory being correct, although this may seem implied by the stress on having empirical relations correspond with theoretical relations. Let us suppose that the theory is wrong but that empirical relations nevertheless correspond with theoretical ones. In that case, there is construct validity. For the sake of drama, we might even imagine a situation where the theory is wrong and the auxiliary assumptions are wrong too, resulting in poor measures, but the falsities cancel each other out to result in correct empirical predictions. In this dramatic scenario, there nevertheless remains construct validity. Thus, at the philosophical level, construct validity seems problematical.

On the other hand, it is possible to argue that false theory and false auxiliary assumptions, though they may, at times, result in empirical relations that match theoretical ones, are unlikely to result in an intricate nomological network supported by many points of intersection with observation. The tacit strategy here is probabilistic: across many studies, using a variety of research paradigms, an intricate pattern of theory-consistent findings is unlikely if the theory is wrong. And the unlikelihood increases if the auxiliary assumptions are wrong too. Thus, the philosophical problem is not necessarily fatal. On the third hand, however, there are cases in the history of science where wrong theories and wrong auxiliary assumptions nevertheless did provide a nomological network that allowed the theory to survive. A case in point would be phlogiston theory, that survived almost two centuries and was finally overturned by Lavoisier [1743-1794], who depended on high quality auxiliary assumptions and extremely precise measurement. Thus, it is possible to have considerable construct validity even when the theory is wrong, and it has happened.

Validity when the theory is wrong also is possible under auxiliary validity, but for a much better reason. Returning to the Geiger counter example, we saw that if the theory is wrong, but the auxiliary assumptions are correct—e.g., the experimental conditions are set up correctly and the Geiger counter works—then the finding (or lack thereof) is

evidence against the wrong theory. The radiation that is supposed to be present is not present, and so there is a good reason to suspect that the theory is false. This seems a relatively straightforward implication and is a point in favor of auxiliary validity.

Moving to the second bullet-pointed question, the answer is again in the affirmative for both construct validity and auxiliary validity, but for different reasons. From the point of view of construct validity, a correct theory is ideal because it helps researchers establish an intricate nomological network. And the goal is aided by having good measures of the constructs in the theory. However, the measures might not be good, in which case empirical and theoretical relations might not correspond, and construct validity will be unimpressive. There is an ambiguity here in that a construct validity perspective does not say why measures might be poor. From the point of view of auxiliary validity, the quality of measures depends on the quality of the underlying auxiliary assumptions. The truth of the theory does not matter for assessing validity (though it obviously matters for other purposes) because the action is in the quality of the auxiliary assumptions. If high-quality auxiliary assumptions are conjoined with a true theory, the results likely will verify the theory; but if high-quality auxiliary assumptions are conjoined with a false theory, the results likely will falsify the theory. Either way is fine: verification implies that the theory has some chance of passing future empirical challenges or of resulting in useful applications whereas falsification implies the opportunity for researchers to make progress by replacing the falsified theory with a better one.

Conclusion

Aviation researchers take as given the desirability of having construct validity, even though many of them misunderstand what it is. The foregoing exposition provides that explanation but also questions whether construct validity should be the emphasis. Thus, there are two main take-home messages.

The first take-home message is that if aviation researchers decide that construct validity should be an important focus, then they should understand what that entails. Establishing a nice factor structure is insufficient. Rather, it is necessary to establish an intricate nomological network that contacts with observations at many points, and that is extremely unlikely to be accomplished in a single study or even a set of studies in a single article. Establishing construct validity is a difficult task and aviation researchers should face, squarely, the extent of the difficulty. Aviation researchers should not view construct validity dichotomously, as having been achieved or not achieved. Rather, they should engage in sober discussions about the extent and intricacy of the nomological network that surrounds the construct and the extent to which it contacts observations. Such discussion would represent a sea change in the aviation literature, but it is necessary if researchers truly care about construct validity as opposed to giving it lip service.

The second take-home message is that construct validity might not be all that it is alleged to be. Although we have seen that some construct validity criticisms can be countered reasonably easily, the criticism featuring auxiliary validity cannot be countered

easily. It is a real problem that measures based on wrong auxiliary assumptions, and therefore lacking in auxiliary validity, nevertheless can result in good construct validity. Thus, researchers might consider shifting their emphasis from construct validity to auxiliary validity.

That said, one reason for the shift—the difficulty of establishing an intricate nomological network, with contact with observations, in the service of construct validity—is insufficient. Auxiliary validity is not necessarily easier to establish than is construct validity. To make the empirical case for auxiliary validity, one would have to have a variety of tests of the auxiliary assumptions underlying the manipulations and measures of concern, a task that researchers are unlikely to complete in a single study. A better reason for the shift concerns philosophical clarity. Researchers who care about auxiliary validity clearly distinguish between constructs in a theory and manipulations and measures, though construct validity enthusiasts might claim to accomplish this too. Although this sort of clarity is crucial, there is another sort of clarity that is crucial too and where the difference in conceptual clarity between the two types of validity manifests clearly.

Under the umbrella of auxiliary validity, there are four categories of possibilities: correct theory and auxiliary assumptions, correct theory and wrong auxiliary assumptions, incorrect theory and correct auxiliary assumptions, and incorrect theory and incorrect auxiliary assumptions. Then, too, there are various gradations involving these. In contrast, under the umbrella of construct validity, one does not distinguish between the theory and the auxiliary assumptions. Rather, researchers support or fail to support theory and measures simultaneously. This lack of clarity may be one reason why construct validity is so poorly understood, despite its dominance in the validity literature since the classic article by Cronbach and Meehl (1955). And the difference in clarity between auxiliary validity and construct validity constitutes a strong reason for preferring auxiliary validity to construct validity. An additional reason for the preference is that auxiliary validity is more easily related to larger philosophy of science issues such as theory verification and falsification (Trafimow, 2020).

Aviation researchers concerned with validity—and they should be concerned with validity—should carefully consider whether to go in the direction of auxiliary or construct validity. I would prefer they choose auxiliary validity. But even if they decide in favor of construct validity, there remains an important message. Aviation researchers should understand that an emphasis on construct validity implies establishing extensive and intricate nomological networks, that contact observations, to simultaneously support the hypothetical constructs of interest, and their measures. Without such nomological networks and corresponding empirical relations, claims to having established construct validity are specious.

References

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Prentice-Hall.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561-571. doi: 10.1001/archpsyc.1961.01710120031004.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Fishbein, M. (1980). Theory of reasoned action: Some applications and implications. In H. Howe & M. Page (Eds.), *Nebraska Symposium on Motivation, 1979* (pp.65-116). University of Nebraska Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Psychology Press (Taylor & Francis).

Fussell, S. G., & Truong, D. (2020). Using virtual reality for dynamic learning: an extended technology acceptance model. *Virtual Reality*. https://doi.org/10.1007/s10055-021-00554-x

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-195). Cambridge, UK: Cambridge University Press.

Lakatos, I. (1978). *The methodology of scientific research programmes: Vol. 1. Philosophical papers*. Cambridge, UK: Cambridge University Press.

Lederman, L. (1993). *The God particle: If the universe is the answer, what is the question?* New York, NY: Houghton Mifflin.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education and national Council on Measurement in Education.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*(1/3), 35–44. http://www.jstor.org/stable/27522333

Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.

Slaney, K. L., & Garcia, D. A. (2015). Constructing psychological objects: The rhetoric of constructs. *Journal of Theoretical and Philosophical Psychology, 35*(4), 244-259. http://dx.doi.org/10.1037/teo0000025

Trafimow, D. (2012). The role of auxiliary assumptions for the validity of manipulations and measures. *Theory & Psychology, 22*(4), 486-498. doi:10.1177/0959354311429996

Trafimow, D. (2017). Implications of an initial empirical victory for the truth of the theory and additional empirical victories. *Philosophical Psychology, 30*(4), 411-433. doi: 10.1080/09515089.2016.1274023

Trafimow, D. (2020). A taxonomy of major premises and implications for falsification and verification. *International Studies in the Philosophy of Science, 33*(4), 211-229, doi: 10.1080/02698595.2021.1964845

Trafimow, D., & Rice, S. (2009). What if social scientists had reviewed great scientific

works of the past? *Perspectives in Psychological Science*, *4*(1), 65–78. doi:10.1111/j.17456924.2009.01107.x