Trends in Psychology
https://doi.org/10.1007/s43076-024-00364-5

ORIGINAL ARTICLE

Building a More Robust Introduction to Measurement in Research Method Courses

Joshua J. Reynolds[1]

Accepted: 22 February 2024

Abstract

Measurement is integral to science. Given that it takes many years to become knowledgeable in measurement, it is valuable to consider current practices in teaching measurement to undergraduate psychology students. It is argued here that psychology research method courses could benefit from significant additions and clarifications in the topic of measurement. Three topics to consider are: discussions of different measurement viewpoints, the conditions for continuous quantities, and measurement challenges in psychology. These topics can be integrated into our courses and would translate to a more nuanced understanding of measurement and a greater ability to critically think about measurement in psychology. Suggested strategies for teaching about these topics are also discussed.

Keywords Measurement · Research Methods · Classical/Realism · Teaching

Measurement is a complex topic and, for a psychology major, the largest dose of measurement content is in the required undergraduate research method course. Many methodology texts for undergraduate psychology students (e.g., Adams & Lawrence, 2019; Cozby & Bates, 2018; Morling, 2021; Nestor & Schutt, 2018; Schweigert, 2021; Stangor, 2014) include, 1) a description of Steven's (1946) four levels of measurement (nominal, ordinal, interval, and ratio), 2) some implicit discussion of operationalism via operational definitions, and 3) a discussion of validity and reliability. Missing is a discussion of other views of measurement, critiques of these different views, fundamental measurement concepts (like conditions for continuous quantities), measurement challenges in psychological science, and discussion as to why psychology has historically been criticized in this area. Given the heavy focus on Steven's scale typology and the missing components, the current topics may not provide students with a significant understanding of measurement. Consequently, this lack of foundational knowledge may lead to questionable measurement

✉ Joshua J. Reynolds
Joshua.Reynolds@scranton.edu

[1] Department of Psychology, University of Scranton, 800 Linden St., Scranton, PA 18510, USA

Published online: 19 March 2024

practices, issues such as validity hacking, being less able to identify measurement issues, and generally less scientifically rigorous research (Flake & Fried, 2020; Hussey & Hughes, 2020).

Instructors may wish to create a more robust discussion and present more fundamental concepts in measurement. While there are many topics that are currently lacking at the undergraduate level, here I discuss adding multiple viewpoints, conditions for continuous quantities, and measurement challenges in psychology. Such topics could be integrated into our courses, which might improve how students understand measurement. In addition, I present some possible strategies for teaching about measurement.

Multiple Viewpoints

A basic understanding of measurement in psychology should include multiple viewpoints. Doing so may promote greater critical thinking by driving students to identify assumptions, evaluate the strengths and weaknesses of different views, synthesize information, and evaluate measurement claims (Bensley et al., 2010). Furthermore, a discussion of multiple viewpoints would honestly represent measurement in research methodology. Viewpoints on measurement include, but are not limited to, classical/realist, operationalist, representationalist, pragmatic, as well as statistical, such as item response theory and Rasch measurement (Maul et al., 2016; Michell, 1997, 1999). These viewpoints are not merely semantic but are fundamentally different views of measurement and numbers.

The realist perspective would define measurement as estimating or discovering continuous quantities (Michell, 1997, 1999). This viewpoint makes a major distinction between simple numerical coding as done when assessing a variable like sex, pain, and even possibly intelligence, as opposed to measurement with a variable like length or mass. Unless that variable has the properties of order and additivity, it cannot be said to be quantitative. So, from the realist perspective, unless there is evidence for quantitative structure of an attribute, then it makes little sense to view the coding of values as measurement. Upon demonstrating quantitative structure, measuring values from an attribute would be equivalent to real numbers (Bostock, 1979). The realist view therefore estimates real numbers. Further, attributes and their structure are the focus.

Operationalism, on the other hand, would call assignment of a number to a construct like sex, measurement. This is seen in Bridgman's (1927, p. 5) argument that "concept is synonymous with the corresponding set of operations", Dingle's (1950, p. 11) argument that measurement is "any precisely specified operation that yields a number", or Steven's (1946, p. 677) reconceptualization of Campbell's (1920) view that measurement is "the assignment of numerals to objects or events according to rule" (Steven's views are also in part representationalist). In psychology, the operationalist viewpoint is typified by Boring's (1945, p. 244) remark that, "…intelligence is what the tests test". From the operationalist viewpoint, meaning is established in empirical operations, and conducting a scientific test involves carrying out an operationally defined measurement (Bickhard, 2001).

In the representationalist view, numbers are assigned to constructs in a way that preserves the qualities of the relations (established through axiomatizations; Krantz et al., 1971; Schwager, 1991). Thus, in the representationalist view, numbers are relatable to real numbers because, given certain conditions, they are isomorphic to real numbers. For example, objects are placed on pans of an equal-arm balance and the relative positions and consequences of removing or adding objects to the pans are observed. The representationalist would view these as empirical operations resulting in empirical relations. Numbers are then assigned or associated in a way that the numerical properties represent the attribute properties. These numbers from a numerical system thus specify some formal relations. This of course rests on the assumption that they are homomorphic or isomorphic. The representationalist therefore assigns numbers, and objects become the focus of measurement.

The view of measurement likely familiar to most psychologists is Steven's (1946) levels of measurement or typology. Steven's (1946) defines measurement as assigning numbers to objects according to rules. Steven's (1946) argued that the rules need to be explicit. After doing so, he derived four different kinds of scale types: nominal, ordinal, interval, and ratio. Further, he argued that given a scale type, only certain kinds of statistical operations were applicable.

Other views of measurement fall under the umbrella of statistics, such as item response theory and Rasch measurement. In the statistical approaches, different mathematical models (Rash model/1PL, 2PL, 3PL) are used to explain the relationship between latent traits and responses (DeMars, 2010). Although understanding these views in detail is likely outside the scope of lower-level undergraduate methods course, instructors could still give a brief description of these approaches and some citations for additional information. They can also be used in other methodology discussions to compare and contrast with classical test theory, which might deepen discussions of concepts such as reliability.

Instructors may also wish to use the conversation of different viewpoints to discuss the distinctions between quantitative and qualitative research. Qualitative research often involves description, experience, and meaning (Hammarberg et al., 2016). Concepts such as validity and reliability can play a role (although there is some disagreement on this, see for example Stenbacka, 2001); however, in qualitative research, some view the researcher as the instrument (Patton, 2002). Thus, measurement is viewed as fundamentally different from the perspective of a qualitative researcher, and this too can enhance student's understanding.

In examining a number of research methods books for psychology and syllabi for method courses (e.g., Adams & Lawrence, 2019; Cozby & Bates, 2018; Morling, 2021; Nestor & Schutt, 2018; Project Syllabus—Society for the Teaching of Psychology, n.d.; Schweigert, 2021; Stangor, 2014), Steven's (1946) typology was found not only to be explicitly discussed, but was presented as the sole view of measurement. In other words, the presentation of Steven's (1946) typology, and only this typology, seems to imply that it is the standard in science. If students are only exposed to Steven's (1946) typology, but not other views, this may create some tunnel vision when thinking about measurement and promote problematic practices. As opposed to only teaching Steven's (1946) typology, I suggest that multiple viewpoints be discussed and criticisms of the perspectives taught. This might promote greater critical thinking and understanding of measurement.

In fact, one of the most problematic views of measurement is Steven's (1946). Velleman and Wilkinson (1993) list over 10 serious deficiencies with the nominal, ordinal, interval, ratio typology. For example, Steven's categories do not describe fixed attributes of data, the categories are not sufficient to describe data scales, and the categories are not exhaustive. To elaborate on the last point, we can easily consider nominal, ordinal, log interval, interval, log discrete interval, discrete interval, ratio, discrete ratio, and absolute (Narens & Luce, 1986). Mosteller and Tukey (1977) discuss names, grades, ranks, counted fractions, counts, amounts, and balances as different levels. There is no justification for deciding that there should be only four.

Furthermore, Steven's (1946) typology and axiomatizations have been argued to be fundamentally flawed since the levels are meaningless, as can be seen when you can follow Steven's (1946) rules and have a variable represented in multiple scale types. Rozeboom (1966) and Prytulak (1975) make this point, as the latter finds that the same scale changes levels depending on how you use it. In discussing this issue, Michell (1990) explains:

> …suppose that a set of people are classified according to eye-color: persons *a*, *b*, and *c* all have blue eyes; *e*, *f*, and *g* all have brown eyes; and both *i* and *j* have green eyes. This classification is an empirical relational system, the relation involved being that of having the same colored eyes as (call it R_1). Let this empirical relational system be numerically represented as follows: assign the number 2 to *a*, *b*, and *c*; the number 3 to *e*, *f*, and *g*; and the number 5 to *i* and *j*. This creates a nominal scale in which the relation of being equal to represents R_1. Now $2 + 3 = 5$ is a numerical relation holding between the numbers used in this nominal scale. (p. 38)

The numbers used to represent eye color (nominal scale) have a numerical relationship, as three and two will add to five. Michell (1990) goes on to explain that inevitably, if the numbers are additively related, then that numerical assignment will also represent a ratio scale:

> $R_2 = \{(a, e, i), (a, e, j), (a, f, i), (a, f, j), (a, g, i), (a, g, j), (b, e, i), (b, e, j)\ (b, f, i), (b, f, j), (b, g, i), (b, g, j), (c, e, i), (c, e, j), (c, f, i), (c, f, j), (c, g, i), (c, g, j)\}$. It represents this relation because R_2 holds between any ordered set of three people *x*, *y*, and *z* (from the original set of eight), if and only if $n(x) + n(y) = n(z)$ (where $n(\mathrm{x})$, $n(\mathrm{y})$, $n(\mathrm{z})$, are the numbers assigned to *x*, *y*, and *z*, respectively). (p. 38-39)

Thus, R_2 can be additively represented. This demonstrates that any nominal scale for some empirical objects also has a corresponding ratio scale with the same numerical assignments. Since for any scale type there will be a set of empirical objects and simultaneously another scale type with that same numerical assignment, to view scale types as being exclusive categories whereupon only certain statistics are permissibly used is problematic.

The statistical point has also been echoed by Velleman and Wilkinson (1993). In fact, Steven's (1951) recognized the issue as well, in writing that technically a standard deviation should not be calculated for an ordinal scale type, which would constitute most of the variables in psychology, but that pragmatically this may be useful. Numerous statisticians such as Tukey (1961) and Guttman (1977) have been critical of the perspective put forward by Steven's (1946), arguing that it is a misuse of statistics to view them in absolute terms, and the same data can be viewed in different ways depending on what the analyst wants to know. Therefore, not only is the typology problematic from a measurement perspective, it is also problematic from a statistical perspective, and might contribute to students incorrectly conducting statistical analyses or thinking they cannot perform a statistical procedure which in reality would be reasonable.

Perhaps most problematically, by conflating measurement and scale types, Steven's (1946) typology obfuscates an important part of measurement, which is understanding the properties of some attribute and how it is structured. Relegating the task of measurement to the assignment of a scale type treats the actual process of measurement as an afterthought.

In the classroom, some of the criticisms against Steven's (1946) typology can also be discussed with regards to the operationalist and representational view more broadly. For example, it is important to understand just how much of a departure the operationalist view of measurement is from the conceptualization of measurement at the time it was introduced and how the operationalist account can lead to arbitrary concept formation (Michell, 1990). The operationalist view of measurement also tends to conflate measurement, meaning, testing, and theory (Bickhard, 2001). Each of these concepts are important and related, but they must not be equated. A consequence of equating them, as done in operationalism, is to hamper progress in any one of them. Indeed, some have argued that the operationalist school of thought has hindered theory development in psychology (Bickhard, 2001). Given the connection with the failed tradition of logical positivism, it is perhaps not surprising that it has many issues (Bickhard, 2001; Vessonen, 2021).

Representationalism also has many issues, such as using objects, rather than attributes, as the focus of measurement. Further, measurement in representationalism always needs to be supplemented by construct-related validation procedures (Michell, 2021; Schwager, 1991). Just like operationalism, in the representationalist view, a claim will be made that an instrument measures something, without any evidence of quantitative structure, thus reducing the importance of a significant scientific process.

Realist accounts can also be critically evaluated. For example, the attributes of inference that are examined in psychology may be unlike those that are studied in physics or chemistry, and requiring the same level of precision is impractical (Guyon et al., 2018; Kyngdon, 2008; Sijtsma, 2012). Michell (1990) has argued for greater use of additive conjoint measurement (see next section) which is consistent with the realist perspective, and also helps to address some of the issues, like lack of control, that are present in studying psychological attributes. However, in practice, it has been more difficult to implement given the limitations of how the axioms are verified and with estimating error (Sherry, 2011; Sijtsma, 2012). Recognizing these issues, those such as Guyon et al. (2018) have argued for a measurement approach in psychology that includes both pragmatism and realism epistemology.

Although not all measurement views were able to be discussed above, and there are many more critiques that could be made, it should be clear that this kind of discussion can be used to greatly enhance students understanding of measurement. Steven's (1946) typology has had a tremendous impact on psychology; thus it is reasonable to still discuss it, but if students are to think critically about measurement, this needs to be done in the context of other views of measurement. Presenting several viewpoints of measurement and their strengths and weaknesses can enhance critical thinking skills, allow students to be better prepared for more advanced measurement concepts encountered at the graduate level, and change the focus to thinking rather than rote memorization of a typology.

Conditions for Continuous Quantities

A fundamental aspect in the history of measurement in science has been the issue of examining if an attribute has quantitative structure. If values from an attribute are quantitative, then they are equivalent to real numbers. This of course has important consequences for using values to make inferences, developing better instruments, and understanding related phenomena. Therefore, it is important to understand what properties are required for quantitative structure. A quantity has the properties of order and additivity. If it also meets conditions for density and completeness, it is continuous.

It is useful to consider length when thinking about these conditions. Hölder (1901) defined the conditions for quantitative structure and showed that length met these conditions. There are three conditions, all of which must be true, for an attribute to have order (Michell, 1990). If X, Y, and Z, are values from some variable, then it must be the case that: 1) if $X \geq Y$ and $Y \geq Z$ then $X \geq Z$ (transitivity); 2) if $X \geq Y$ and $Y \geq X$ then $X = Y$ (antisymmetry); and 3) either $X \geq Y$ or $Y \geq X$ (strong connexity). Additivity has six conditions, all of which must be true: 1) $X + (Y + Z) = (X + Y) + Z$ (associativity); 2) $X + Y = Y + X$ (commutativity); 3) $X \geq Y$ if and only if $X + Z \geq Y + Z$ (monotonicity); 4) if $X > Y$ then there exists a value Z such that $X = Y + Z$ (solvability); 5) $X + Y > X$ (positivity); and 6) there exists a natural number n such that $nX > Y$ (Archimedean condition). If some attribute met all nine conditions (i.e., has order and additivity), it can be considered a quantity. Additionally, there exists conditions for being continuous. If all the previous conditions are true and it is also true that: 1) if X and Y are any values of Q such that $X > Y$ then there exists Z in Q such that $X > Z > Y$ (density) and 2) every non-empty set of values of Q which has an upper bound has a least upper bound (least-upper-bound property/completeness), then the variable is a continuous quantity.

It is important to understand that the values X, Y, and Z denote specific magnitudes of some attribute, like length, independent of any actual physical measurement. Regardless of whether or not I measure the length of my desk, the desk has the attribute length. Similarly, if $X + Y = Z$, this means that length Z is made of discrete parts X and Y, outside of any human operation (Michell, 2003). Notice that these conditions are not akin to *how* a researcher demonstrates quantitative structure. The conditions are important because they make explicit what we are inferring when we claim quantitative structure for an attribute. This author argues that this should be the key aspect of the discussion. Instructors can use examples like length, volume, and mass to point out that we correctly make the kinds of inferences with those attributes, commensurate with their status as having quantitative structure. Then, an attribute like intelligence, can be used as a contrast to discuss the idea that intelligence likely has order, but not necessarily additivity. Thus, the inferences that we draw regarding values from an intelligence instrument are different than if it was length, for example.

It should be noted that while undergraduate method books may make the distinction between a discrete value and a continuous value, or a category and a number, this is often described in very little detail and not tied explicitly to the structure of some variable, which is at the core of measurement. Instead, the distinction is more often erroneously described as a decision the researcher must make.

In this discussion of what makes for quantitative structure, another important distinction that must be made is that between extensive and intensive variables (Landsberg, 1978; Michell, 1990; Redlich, 1970). A variable like mass is extensive. Extensive properties depend on the amount of matter in a sample. A larger sample of lead means more mass. Other common examples include length, volume, entropy, and energy. A variable like temperature is intensive. Intensive properties depend on the type of matter, not the amount. For example, taking two glasses of milk each at 40° F and putting them into one large container does not make the temperature 80° F. Other intensive examples include solubility, density, and hardness. The key to this discussion is understanding that variables like length and temperature are both quantities, as they are both ordered and additive, but they do not demonstrate these properties in the same manner. Extensive variables like mass and length can demonstrate their order and additivity by concatenation (joined side-by-side). On the other hand, intensive variables like temperature, are not concatenated.

Despite length being extensive, it was still a challenge to show that it met the conditions for a continuous quantity (Hölder, 1901). Understanding the properties of intensive variables, like temperature, is even more challenging (Barnett, 1956; Chang, 2004; Sherry, 2011). This is quite relevant for psychology, as the kinds of variables that are examined in our field are more like intensive variables, as they cannot be concatenated (e.g., self-esteem). Therefore, to understand measurement in psychology, it is important for students to understand not only that attributes can have different properties like order and additivity, but that they can demonstrate these properties in fundamentally different ways.

It can be difficult to uncover the properties of an attribute, particularly if that attribute is a psychological one. As discussed more in the next section, it can also be challenging when the methodologist lacks precise control of the variables, and the variables are dependent on the environmental context. This is partly why the development of additive conjoint measurement (ACM) marks significant progress. ACM allows us to bypass some of these issues and test for quantitative structure.

In ACM, at least two attributes, A and X, non-interactively (e.g., additively, multiplicatively, etc.), relate to a third attribute, P (e.g., volume, mass, and density; Luce

& Tukey, 1964; Luce et al., 1990; Michell, 1990). It is required that P possesses an infinite number of values, P = *f*(A, X) where *f* is some mathematical function, there is a simple order, ≥ upon the values of P, and values of A and X can be identified (that is, objects can be classified according to the value of A and X they possess). If three conditions hold [double cancellation (described later), solvability, and the Archimedean condition], then, P, A, and X are quantitative and *f* is a noninteractive function. Importantly, it is not a requirement that P, A, or X, are previously known to be continuous quantities, nor is it a requirement that the variables be concatenable. Thus, ACM is particularly useful in quantifying variables in cases where it is known that the variables have order but they cannot be concatenated. This is often the situation in psychology.

One can use ACM to show evidence that some attribute (e.g., P, the probability of getting an item correct) is related additively to two others (e.g., A the ability and X the item difficulty) such that P = *f* (A + X) (where *f* is any positive monotonic function; Heene, 2013). If a psychologist hypothesized that trait self-control has quantitative structure, it is possible to use ACM to validate its structure, despite the difficulty in precisely controlling values of self-control. From a measurement perspective, the development of ACM is a significant step forward, and it can even be shown that the conditions for continuous quantities are a special case. As discussed above, however, there are also practical limitations to ACM which has contributed to it not receiving more widespread use (Heene, 2013; Sijtsma, 2012). Some also view the conditions in ACM as still too restrictive (Maul et al., 2016). Despite some limits of ACM, undergraduate students could benefit greatly by being exposed to measurement approaches that are specifically useful in psychology that allow us to *test* for quantitative structure.

Challenges for Psychology

Although researchers may hypothesize that some psychological attribute, like intelligence, has quantitative structure, it may be that it has order but not additivity, and therefore is not quantitative. Investigating structure of psychological attributes is a monumental task given that many of the attributes studied in psychology cannot be concatenated, tend to be highly socially dependent, and can have properties that emerge from micro-elements and processes that the macro-element cannot be reduced to (Guyon et al., 2018; Humphreys, 2008). No doubt most research methods courses in psychology discuss some of the challenges of studying psychological attributes. Challenges that could receive more widespread discussion in our courses include separation, manipulability, and invariance, discussed below. By teaching these challenges to students, we enhance critical thinking and better prepare students for understanding and conducting scientific research.

When a thermometer is used to record values of temperature, we are not directly measuring temperature. We are observing the effect that temperature has on the volume of a substance sealed in a glass container. So, there is some separation from what we have (readings on a glass tube about volume of a liquid) and what we want (temperature; Bond et al., 2021). Separation can be seen when values of an instrument are indirectly related to the variable under investigation.[1] Despite that with temperature and thermometers, there is only a small degree of separation, the development of accurate thermometers still took several centuries (Barnett, 1956; Chang, 2004; Sherry, 2011). In the case of intelligence, a value recorded on an instrument like the WAIS or Ravens progressive matrices is not what we want; rather we want to know the value of some unobserved variable (Neisser et al., 1996). For example, we want to know Person A's intelligence, but since the test score is indirectly a product of the variable of interest, a WAIS score of 112 is separated from their true level of intelligence. As an example, suppose intelligence was akin to a particular network of neurons in the brain. This network would likely send its output to dozens of other neural systems, which in turn would send their output to dozens of other systems (Fingelkurts & Fingelkurts, 2004). Thus, the separation observed between the state of this intelligence network and the eventual score that is produced on the WAIS means more deviations and thus more noise. The same would be true for many other instruments for constructs like depression, self-control, hunger, etc.

The issue of separation is particularly relevant in dynamic systems, like the human body. A hallmark of dynamic systems is sensitivity dependence on initial conditions (Lorenz, 1963; Ruelle, 1991). In essence, a small initial difference can cause a change in what happens later. Thus, even in a completely deterministic system, a very small initial difference can result in no relationship being found between the initial state of the system and the path the system takes later on. When making inferences about some unobserved variable, like intelligence, a small difference introduced early on in the network would theoretically result in a different score on the test.

The issue of separation can be addressed by developing a precise understanding of the causal process involved in the system under investigation as well as creating instruments or techniques that can make more direct inferences on the attributes of interest. In the history of temperature, concepts of capacity for heat/specific heat and latent heat, as well as an understanding of the causal processes involved, facilitated the development of temperature instruments (Barnett, 1956; Bringmann & Eronen, 2016; Chang, 2004; Sherry, 2011). At present, psychology does not have a precise understanding of how genes, environments, neurons, etc. causally work to produce scores on a test, and separation is arguably an issue here.

A second challenge is one of manipulability. In physics or chemistry, many times the variables of interest can either be directly or indirectly manipulated. The ability to precisely manipulate a variable and carefully observe the effects has been essential in measurement in the history of science. For example, being able to add a specified known quantity of heat and observe the specific change on a thermometer facilitated the development of thermometers (Barnett, 1956; Chang, 2004; Sherry, 2011).

[1] This is conceptually related to the manifest vs. latent variable distinction. A latent variable might be somewhat separated from the values on an instrument, or it might be many orders separated. Thus, separation is not just that a variable is not directly observed, but how directly the measure relates to the variable. The argument presented here is that the more separated, the more difficult to study and develop instruments for.

Further, being able to manipulate pressure and show that readings on some instruments were affected was important in the development of thermometers. In some cases, psychological variables cannot be manipulated because of ethical issues (e.g., growing up in a high crime neighborhood). In other cases, it would be impractical (e.g., the number of friends you have). In still other cases, it is not even clear what it would mean to manipulate a variable (e.g., intelligence). If it is difficult to take one unit of intelligence in one condition and two units of intelligence in another condition, and test how an instrument like the WAIS responds, it becomes more difficult to develop precise measurements. Therefore, the lack of the ability to manipulate many psychological variables poses a challenge, and students learning psychological science should recognize this challenge to understand measurement.

A third challenge is invariance. Invariance is the concept that values attributed to variables by a measurement system should be independent of a particular instrument. In other words, “you can’t measure change with a measure that changes” (Bond et al., 2021, p. 69). If our psychological variables themselves do not exist independent of our particular cultural, social, and historical conditions, the instruments will not be invariant. Invariance is an important aspect of measurement. When we conceptualize values of length, mass, temperature, etc., they exist independent of the place and time. But many psychological variables either likely or definitively do not exist independent of place and time (Hussey & Hughes, 2020). For example, what self-esteem means is at least partly dependent on particular cultural, social, and historical conditions (Baumeister & Leary, 1995; Howell et al., 2019; Maslow, 1943). Suppose the average levels of self-esteem were compared between two countries. Country A has an average of 3 and country B has an average of 4. If the measurement is not invariant, it is not clear that country B has higher average self-esteem. A score of 4 on the scale in country B, might equate to a 2 in country A. The issue is invariance. It could be the case that within some socially dependent conditions, values of a self-esteem assessment are invariant, but this would be contextually dependent, and we would treat that measurement system differently. Some instruments for assessing intelligence, for example, are frequently renormed, and we would not necessarily equate a score of 100 today to a score of 100, 50 years ago (Flynn, 1987).

How to demonstrate invariance is important in measurement. This could be an opportunity in a course to link issues of measurement in psychology with some of the different viewpoints of measurement, namely item response theory and Rasch measurement. Differential item functioning is used in Rasch measurement to examine if groups (e.g., different ages or ethnicities) have different probabilities of endorsing a given item on a scale, after controlling for overall scores. This feeds back to a discussion of differing viewpoints on measurement and gives an opportunity for interleaving topics, which improve learning and memory (Rohrer, 2012).

Suggested Strategies for Teaching

While there was some discussion of teaching strategies embedded into the discussion above, specific strategies are explicated here. The suggested strategies are meant only to give instructors more ideas on how to incorporate the previously discussed material. Whether the proposed strategies discussed here are effective is an empirical question, but one beyond the scope of the current work.

One potential strategy to help students understand measurement is to introduce to them different variables with different properties and have them discover the nature of their structure for themselves. For example, an instructor might briefly define length (i.e., the straight-line distance between two points along an object) and sex (a biological quality based on the type of gametes produced), and then have students in small groups try to determine why they are different and what properties the variables have. In such an exercise, students might even discover the properties of order and additivity themselves.

Another feasible strategy, specifically for teaching different views of measurement, could be to present statements and then examine how each viewpoint would think about them. For example, a statement like, "We can use a ruler to measure the length of a hand" would be an acceptable statement from the realist, operationalist, and representationalist view; however, a statement like "We can use a self-esteem scale to measure self-esteem", is unacceptable from the realist perspective because there is little evidence that self-esteem is a quantity. It would also be possible to bring up criticisms of traditional self-esteem scales from an item response and Rasch measurement perspective.

Another strategy could be to use temperature as an example to tie in several concepts. As mentioned above, most variables do not manifest their additivity by concatenation; variables like length and mass are the exceptions, not the rule. More often, variables are like temperature, which is an excellent example of an intensive attribute. While the ability to construct accurate measures of temperature took several centuries, scientists had explicitly or implicitly hypothesized that temperature was quantitative (Sherry, 2011). There are many variables in psychology that are hypothesized to be quantitative, like intelligence. It is instructive to explain some of the measurement challenges that scientists faced in the history of temperature. For example, Galileo's 1592 device was technically a barothermoscope, as it was affected by not just temperature, but air pressure. In creating modern thermometers, many liquids were tried, including mercury, which has the advantage of increasing volumetrically linear over a wide range of values and being less liable to cling to the glass tube compared to alcohol. The thermometer was refined over centuries and the physical systems that affect temperature measurements are relatively simple, compared to human behavior. Through a discussion of temperature, students might gain an appreciation for how serious and arduous the measurement process is. Temperature is also an excellent example where the measurement of the variable was greatly aided by development of theory and understanding of the relevant causal processes (Bringmann & Eronen, 2016). Therefore, temperature can be used to show what psychology is often lacking and why measurement in psychology is challenging. Lastly, temperature is a very useful example to show how we can have readings from an instrument that relates to something we do not want (e.g., volume of mercury), and use those readings to make an inference to something we do want (e.g., value of temperature). In psychology, we may have scores on a test or answers on a questionnaire, what we don't want, and we would like to make an inference about some hypothesized latent variable (Bond et al., 2021).

Another strategy that may be useful in teaching measurement in psychology is to teach history. At the very beginning of psychology, there were debates as to the scientific status of the field and debates about keys issues, such as what are we actually measuring in psychology. The reason measurement in psychology was questioned concerned how measurement was and is viewed in the natural sciences. A classical/realist viewpoint of measurement suggests that in many cases, psychologists were not measuring, but rather numerically coding. In many ways this came to a head with the 1932 committee established by the British Association for the Advancement of Science. The committee focused on psychometrics, some of the most systematic and precise research done in psychology, including Steven's sone scale, and concluded that what was being called measurement was not. This called into question the entire field of psychology. Michell (1990) argued that the reaction of psychologists in the early to mid-1900s is what lead to the adoption of a view of measurement that departs from the natural sciences. Indeed, it does appear that ideas like Steven's (1946) typology were embraced because it legitimatized the practices of psychologists as measurement. Unfortunately, psychology texts rarely if ever mention this history. Not only is even a brief discussion of this history more intellectually honest, but it also puts the differing viewpoints of measurement and conditions for quantities in context, which may facilitate understanding and critical thinking.

Lastly, we might consider some teaching strategies for conditions of quantities. The focus of the conditions is to make it clear that when we are claiming quantitative structure, those conditions are what we are inferring to be true. Most of the conditions are relatively straightforward, and students may already be familiar with some of them, such as transitivity. Examples of variables that meet or do not meet the different conditions can be useful, such as sex, which has neither order nor additivity, to mass, which has order, additivity, density, and completeness. Students may initially struggle with the Archimedean condition, which is also part of ACM (Luce et al., 1990; Michell, 1990). While the Archimedean condition can be written many ways, simply stated, it means that we can take any two positive numbers, X and Y, and find a natural number such that nX is larger than Y. Or similarly, some set of natural numbers is not bounded above. As way of example, we could imagine trying to fill-up a bathtub with salt. No matter how small the spoon, as long as we have enough salt, there is some natural number of spoonful's that will allow us to overflow the bathtub with salt. This is often also illustrated as in Fig. 1. By way of analogy and simple figures, it might be easier to understand an abstract concept like the Archimedean condition.

If ACM was introduced, double cancellation might also be somewhat difficult. As often stated, with a, b, and c being values of A, and x, y, and z being values of X, and this being related to a noninteractive relationship with P, then the following must be satisfied: if ay ≥ bx, and bz ≥ cy, then az ≥ cx (Luce et al., 1990; Michell, 1990). This is actually straightforward if we understand double cancellation as a more complex form of transitivity. Simply stated, it concerns *pairs* of values of P, ordered by ≥, which relates to order in other particular *pairs* of values. While there is much to be said about double cancellation, the condition can be more easily grasped by a schematic representation. If a_1:a_3 are 3 values of A and x_1:x_3 are 3 values of X, they can be grouped in a matrix, as in Fig. 2, and using arrows to represent a ≥ relationship. If

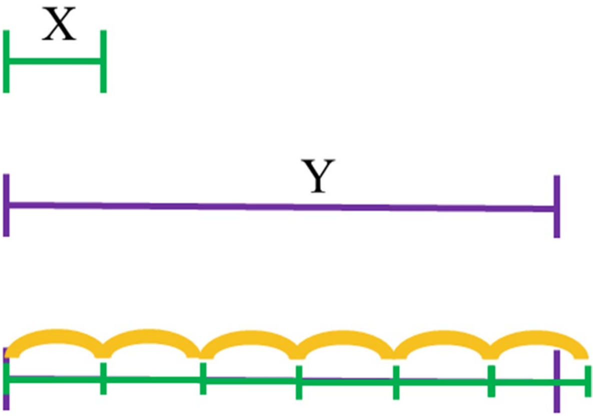
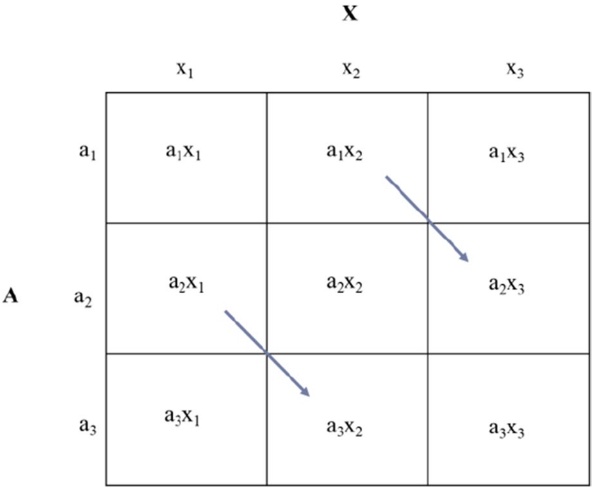


Fig. 1 Illustration of the Archimedean condition



		X		
		x_1	x_2	x_3
A	a_1	a_1x_1	a_1x_2	a_1x_3
	a_2	a_2x_1	a_2x_2	a_2x_3
	a_3	a_3x_1	a_3x_2	a_3x_3

Fig. 2 Illustration of using matrices to discuss double cancelation

double cancellation holds, a_1x_2 should be greater than or equal to a_2x_3 and a_2x_1 should be greater than or equal to $a_3x_{2,}$ etc. Michell (1990) uses matrices to illustrate the 36 different substitution instances from this example, and the relationships which must be true (see also Perline et al., 1979, for a discussion of how this relates to the Rasch model). Given that this would be meant for a discussion in an undergraduate research method course, it may be unnecessary to give a complete account of conditions like double cancellation or the Archimedean property. Nonetheless, some brief examples like this might help students. A more advanced course could then explicate how scientists in practice can use ACM, for example, to examine quantitative structure or how this is connected to Rasch measurement.

Reflections and Conclusions

Many of the ideas discussed here have been passionately expressed in previous work. Rozeboom (1966), for example, not one to mince words, discussed why he viewed Steven's (1946) typology as "complete nonsense" (p. 188). Although *methodologists* are familiar with the issues, the dominant view in undergraduate methodology textbooks is still Steven's (1946), which suggests that the concepts described here are not commonly taught in undergraduate (and possibly graduate) courses. A cause for this could be that works like Rozeboom (1966) are aimed more at methodologists and psychometricians, than instructors. This work is intended to, in part, fill that gap by bringing these issues to the attention of methodology instructors and providing some options for them to implement the suggestions.

A consequence for the ubiquity of Steven's (1946) views in undergraduate psychology books is that it may cause people to believe that it is true. This frequent presentation could be triggering an illusory truth effect (Hasher et al., 1977; Hassan & Barber, 2021). Further, some might object to the suggestions here and argue that the ubiquity of Steven's (1946) typology suggests that is the correct approach. Yet this commits the appeal to common practice fallacy. Just because it is common does not make it true (Lee, 2017). Steven's (1946) typology is easy to understand and comes with a built-in mnemonic, NOIR, making it easy to understand compared to some of the suggested material here, like ACM. However, this processing fluency may unfortunately cause people to believe that it is true (Alter & Oppenheimer, 2006). Similarly, given familiarity over time, we might even feel positive about the typology, a possible mere exposure effect (Bornstein, 1989). Thus, despite the problems with Steven's (1946) typology, students and instructors alike could develop a preference for it, and this may contribute to why it is still taught.

Research suggests that need for cognition and previously stored knowledge do not prevent an illusory truth effect (Fazio et al., 2015; Newman et al., 2020). On the other hand, some evidence suggests that having people act as "fact checkers" and evaluate claims by focusing on accuracy at exposure can prevent it (Brashier et al., 2020). The teaching strategy of having students think about the properties of different variables for themselves and exposing students to multiple viewpoints and having them discuss the strengths and weaknesses in each view is argued to facilitate this kind of critical thinking. Furthermore, this problem-solving-then-instruction tactic might give students opportunities to notice and encode key features of measurement on their own, and is consistent with productive failure (Sinha & Kapur, 2021).

Psychology texts currently focus heavily on Steven's (1946) typology and without support from the required textbook, some instructors may understandably be hesitant in incorporating some of these suggestions. The teaching strategies discussed above may help, but there are also several excellent articles that could be assigned, like Michell (2003), which gives a brief introduction to measurement in science and covers several of the topics discussed here. Authors of undergraduate method texts are encouraged to incorporate some of suggested content in the future, which would help instructors.

A reasonable objection to the recommendations here might be that this material, while important, is too difficult for undergraduate students, and might be better suited to graduate courses. There are at least three responses to such an argument. First, what is advocated here amounts to an *introduction* to these issues. These topics can be further explicated in more advanced undergraduate courses or graduate courses. Second, most students will not attend graduate school and therefore most psychology majors will continue to have a narrow view of measurement unless undergraduate curriculum is updated. Third, measurement discussions, by their nature, are complex. Some students may indeed struggle, but just as not all students can succeed in fields like physics or engineering, some students may not be suited to a rigorous scientific examination of psychological methodology. The justification for content inclusion is arguably established by determining what is essential, not necessarily what is easy.

Another possible argument could be that, in terms of different views of measurement, the debate is just semantics. As discussed above, the different views represent different ways of not just discussing measurement but investigating empirical phenomenon. Science works to understand the structure of empirical phenomena and views such as operationalism equivocate assignment of numerals to the scientific task of measurement (Michell, 1990). Thus, an operationalist view tends to disregard the investigation of structure. Representationalism investigations must also be supplemented with construct validation procedures that may be inadequate for the development of robust instruments as seen in other scientific disciplines (Maul, 2017; Michell, 2021). Therefore, these different views are not merely semantic arguments, but have genuine consequences for how research is conducted.

As instructors, we must continuously reflect and question if our current approach is optimal or if there are opportunities for improvement. Given the above discussion, there are gaps in the standard curriculum and thus opportunities. The arguments made here should not be regarded as a definitive set of suggested content. They are merely elements instructors could consider incorporating to build a more robust introduction to measurement. If only one of the elements could be integrated, the different viewpoints of measurement could be considered the most important. Indeed, issues in measurement are to be discussed and debated, not rote memorized. Although the pragmatic view, item response theory, and Rasch measurement were not examined in detail in the above discussion, students would also benefit from a basic introduction to those views. Guyon et al. (2018) discuss the utility of a pragmatism-realism viewpoint in psychology, which might be an appropriate reading for students. Similarly, Bailes and Nandakumar (2020) give an excellent introduction to Rasch measurement, how it compares to classical test theory approaches, and the practical aspect of conducting Rasch analyses. Lastly, Nguyen et al. (2014) discuss the basic aspects of item response theory and present an application for patient-reported outcome measures. DeMars (2010) is also an excellent short text on item response theory.

By incorporating a more rigorous presentation of measurement at the undergraduate level, this may translate to a more nuanced understanding of scientific research and eventually, shrewder scholarship. Although the focus of this discussion has been for research method courses, the measurement concepts can and should be incorporated in other psychology courses when relevant. Hopefully, this paper offers a starting point for those who want to add a more robust discussion of measurement to their classes.

Funding There is no funding source.

Data Availability (data transparency) Not applicable.

Code Availability (software application or custom code) Not applicable.

Declarations

Informed Consent Not applicable.

Conflict of Interest The author declares no conflict of interest.

References

Adams, K. A., & Lawrence, E. V. (2019). *Research methods, statistics, and applications* (2nd ed.). Sage Publications.

Alter, A.L., & Oppenheimer, D.M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences*, *103*, 9369-9372. https://doi.org/10.1073/pnas.0601071103

Bailes, L. P., & Nandakumar, R. (2020). Get the most from your survey: An application of Rasch analysis for education leaders. *International Journal of Education Policy and Leadership*, *16*(2). https://doi.org/10.22230/ijepl.2020v16n2a857

Barnett, M. (1956). The development of thermometry and the temperature concept. *Osiris, 12*, 269–341.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bensley, D. A., Crowe, D. S., Bernhardt, P., Buckner, C., & Allman, A. L. (2010). Teaching and assessing critical thinking skills for argument analysis in psychology. *Teaching of Psychology, 37*(2), 91–96. https://doi.org/10.1080/00986281003626656

Bickhard, M. H. (2001). The tragedy of operationalism. *Theory & Psychology, 11*(1), 35–44. https://doi.org/10.1177/0959354301111002

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model. Fundamental measurement in the human sciences* (4th ed.). Routledge.

Boring, E. G. (1945). The use of operational definitions in science. *Psychological Review, 52*, 243–245.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin, 106*(2), 265–289. https://doi.org/10.1037/0033-2909.106.2.265

Bostock, D. (1979). *Logic and arithmetic* (Vol. 2). Clarendon Press.

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition, 194*, 104054. https://doi.org/10.1016/j.cognition.2019.104054

Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.

Bringmann, L. F., & Eronen, M. I. (2016). Heating up the measurement debate: What psychologists can learn from the history of physics. *Theory & Psychology, 26*(1), 27–43. https://doi.org/10.1177/0959354315617253

Campbell, N. R. (1920). *Physics: The Elements*. Cambridge University Press.

Chang, H. (2004). *Inventing temperature*. Oxford University.

Cozby, P. C., & Bates, S. (2018). *Methods in behavioral research* (13th ed.). McGraw-Hill.

DeMars, C. (2010). *Item response theory*. Oxford University Press.

Dingle, H. (1950). A theory of measurement. *British Journal of the Philosophy of Science, 1*, 5–26.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002. https://doi.org/10.1037/xge0000098.supp

Fingelkurts, A. A., & Fingelkurts, A. A. (2004). Making complexity simpler: Multivariability and metastability in the brain. *International Journal of Neuroscience, 114*(7), 843–862. https://doi.org/10.1080/00207450490450046

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *PsychologicalBulletin, 101*(2), 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Guttman, L. (1977). What is not what in statistics. *The Statistician, 26*, 81–1107.

Guyon, H., Kop, J. L., Juhel, J., & Falissard, B. (2018). Measurement, ontology, and epistemology: Psychology needs pragmatism-realism. *Theory & Psychology, 28*(2), 149–171. https://doi.org/10.1177/0959354318761606

Hammarberg, K., Kirkman, M., & de Lacey, S. (2016). Qualitative research methods: When to use them and how to judge them. *Human Reproduction, 31*(3), 498–501. https://doi.org/10.1093/humrep/dev334

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior, 16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive Research, 6*(1), 1–12. https://doi.org/10.1186/s41235-021-00301-5

Heene, M. (2013). Additive conjoint measurement and the resistance toward falsifiability in psychology. *Frontiers in Psychology, 4*, 246. https://doi.org/10.3389/fpsyg.2013.00246

Hölder, O. (1901). Die Axiome der Quantitat und die Lehre vom mass [The axioms of quantity and the theory of mass]. *Sachsische Akademie Wissenschaften Zu Leipzig, Mathematisch-Physische Klasse, 53*, 1–64.

Howell, J. L., Sosa, N., & Osborn, H. J. (2019). Self-esteem as a monitor of fundamental psychological need satisfaction. *Social and Personality Psychology Compass, 13*(8), e12492. https://doi.org/10.1111/spc3.12492

Humphreys, P. (2008). Computational and conceptual emergence. *Philosophy of Science, 75*(5), 584–594. https://doi.org/10.1086/596776

Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 166–184. https://doi.org/10.1177/2515245919882903

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Additive and polynomial representations* (Vol. 1). Academic Press.

Kyngdon, A. (2008). The Rasch model from the perspective of the representational theory of measurement. *Theory & Psychology, 18*, 89–109. https://doi.org/10.1177/0959354307086924

Landsberg, P. T. (1978). *Thermodynamics and statistical mechanics*. Oxford University.

Lee, S. F. (2017). *Logic: A comprehensive introduction*. Hodder & Stoughton.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Science, 20*, 130–141.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement: Representation, axiomatization, and invariance* (Vol. 3). Academic Press.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1*(1), 1–27.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396. https://doi.org/10.1037/h0054346

Maul, A., Torres Irribarra, D., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement, 79*, 311–320. https://doi.org/10.1016/j.measurement.2015.11.001

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives, 15*(2), 51–69. https://doi.org/10.1080/15366367.2017.1348108

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Routledge.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355–383. https://doi.org/10.1111/j.2044-8295.1997.tb02641.x

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge University Press.

Michell, J. (2003). Measurement: A beginner's guide. *Journal of Applied Measurement, 4*(4), 298–308.

Michell, J. (2021). Representational measurement theory: Is its number up? *Theory & Psychology, 31*(1), 3–23. https://doi.org/10.1177/0959354320930817

Morling, B. (2021). *Research methods in psychology: Evaluating a world of information* (4th ed.). W. W. Norton & Company.

Mosteller, F., & Tukey, J. W. (1977). *Data Analysis and Regression*. Addison-Wesley.

Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin, 99*(2), 166–180. https://doi.org/10.1037/0033-2909.99.2.166

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77–101. https://doi.org/10.1037/0003066X.51.2.77

Nestor, P. G., & Schutt, R. K. (2018). *Research methods in psychology: Investigating human behavior* (3rd ed.). Sage Publications.

Newman, E. J., Jalbert, M. C., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition: An International Journal*, *78*. https://doi.org/10.1016/j.concog.2019.102866

Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient, 7*(1), 23–35. https://doi.org/10.1007/s40271-013-0041-0

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Sage.

Prytulak, L. S. (1975). Critique of S.S. Stevens' theory of measurement scale classification. *Perceptual and Motor Skills, 41*(1), 3–28. https://doi.org/10.2466/pms.1975.41.1.3

Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement, 3*(2), 237–255. https://doi.org/10.1177/014662167900300213

Project Syllabus - Society for the Teaching of Psychology (n.d.). Retrieved from https://teachpsych.org/otrp/syllabi/index.php/#methods

Redlich, O. (1970). Intensive and extensive properties. *Journal of Chemical Education, 47*, 154–156. https://doi.org/10.1021/ed047p154.2

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

Rozeboom, W. W. (1966). Scaling theory and the nature of measurement. *Synthese, 16*, 170–233.

Ruelle, D. (1991). *Chance and Chaos*. Princeton University Press.

Schwager, K. W. (1991). The representational theory of measurement: An assessment. *Psychological Bulletin, 110*(3), 618–626. https://doi.org/10.1037/0033-2909.110.3.618

Schweigert, W. A. (2021). *Research methods in psychology: A handbook* (4th ed.). Waveland Press.

Sherry, D. (2011). Thermoscopes, thermometers, and the foundations of measurement. *Studies in History & Philosophy of Science Part A, 42*(4), 509–524. https://doi.org/10.1016/j.shpsa.2011.07.001

Sijtsma, K. (2012). Psychological measurement between physics and statistics. *Theory & Psychology, 22*, 786–809. https://doi.org/10.1177/0959354312454353

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798. https://doi.org/10.3102/00346543211019105

Stangor, C. (2014). *Research methods for the behavioral sciences* (5th ed.). Cengage Learning.

Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. *Management Decision, 39*(7), 551–555. https://doi.org/10.1108/EUM0000000005801

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680. https://doi.org/10.1126/science.103.2684.677

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). Wiley.

Tukey, J. W. (1961). Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments, in The Collected Works of John W. Tukey, Vol. III (1986), ed. L. V. Jones, Wadsworth, pp. 391–484.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician, 47*(1), 65–72. https://doi.org/10.2307/2684788

Vessonen, E. (2021). Conceptual engineering and operationalism in psychology. *Synthese, 199*, 10615–10637. https://doi.org/10.1007/s11229-021-03261-x