

Squaring the circle: From latent variables to theory-based measurement

Theory & Psychology
2023, Vol. 33(1) 118–137
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09593543221127985
journals.sagepub.com/home/tap



Matthias Borgstede 

University of Bamberg

Frank Eggert 

Technische Universität Braunschweig

Abstract

Psychometrics builds on the fundamental premise that psychological attributes are unobservable and need to be inferred from observable behavior. Consequently, psychometric procedures consist primarily in applying latent variable modeling, which statistically relates latent variables to manifest variables. However, latent variable modeling falls short of providing a theoretically sound definition of psychological attributes. Whereas in a pragmatic interpretation of latent variable modeling latent variables cannot represent psychological attributes at all, a realist interpretation of latent variable modeling implies that latent variables are empty placeholders for unknown attributes. The authors argue that psychological attributes can only be identified if they are defined within the context of substantive formal theory. Building on the structuralist view of scientific theories, they show that any successful application of such a theory necessarily produces specific values for the theoretical terms that are defined within the theory. Therefore, substantive formal theory is both necessary and sufficient for psychological measurement.

Keywords

formal models, latent variables, measurement, metatheoretical structuralism, theory-based measurement

Psychological measurement is a controversial topic. On the one hand, there are many well-established psychometric procedures that aim to provide psychological measurements. These procedures usually consist in the assessment of an individual's behavior in a standardized test situation (e.g., the answers to a collection of verbal statements or the comparison

Corresponding author:

Matthias Borgstede, University of Bamberg, Markusplatz 3, Bamberg, D-96047, Germany.
Email: matthias.borgstede@uni-bamberg.de

and rating of certain objects), which is then transformed into one or more numerical variables. On the other hand, the epistemological and ontological status of such *measurements* has been repeatedly questioned. The critique ranges from the untested (and possibly wrong) assumptions implicit in psychometric models, and inconsistencies in their underlying conceptual framework, to semantic vagueness with regard to psychological concepts.

One of the most prominent issues discussed in this context is that the present *modus operandi* builds on several untested assumptions about the nature of psychological attributes and their relation with the behavior that is actually assessed. For example, Michell (1997, 1999, 2009) argues that the quantitative nature of psychological attributes cannot be taken for granted but requires empirical justification. In his view, the question of whether psychological attributes are quantitative can only be justified if the empirical structure of psychological assessments conforms to the axioms of a corresponding measurement model (Krantz et al., 1971). For example, the quantitative nature of an attribute such as subjective brightness needs to be demonstrated by showing that the differences on the numerical scale produced by psychological assessment methods actually represent differences in the underlying attribute. There are many more implicit assumptions underlying psychometric procedures. For example, the common practice of ignoring the dynamic processes that generate behavior requires that the underlying dynamic system is ergodic (i.e., that the assessed behavior is not history-dependent; Mangalam & Kelty-Stephen, 2021; Molenaar, 2008; Olthof et al., 2020). A similar problem arises when attributes that vary between individuals are illegitimately equated with attributes that vary within individuals (Fisher et al., 2018; Molenaar, 2004).

In addition to the debate about the ontology and epistemology of psychological attributes, the present *modus operandi* has been criticized because it builds on an inconsistent conceptual framework (Maraun & Gabriel, 2013). The numerical variables resulting from psychological assessment procedures are often equated with the underlying attributes one wishes to measure. These, in turn, are routinely conflated with the concepts that refer to them. For example, the concept of *intelligence* may refer to a hypothetical psychological attribute, which in turn might be representable by a numerical variable. Whereas variables can be consistently described as special kinds of (quantitative) concepts (Borgstede & Scholz, 2021), they are definitely to be distinguished from the hypothesized attributes to which they refer (Uher, 2021). Therefore, it would be wrong to state that one “measures a concept” or that “a variable is hypothetical.” This kind of category mistake is easily overlooked, since all three terms (concept, attribute, and variable) are often subsumed under the term *construct* (Slaney & Garcia, 2015).

The conceptual ambiguity within psychometrics becomes especially problematic when one considers that the proposed psychological concepts are themselves inherently vague (Flake et al., 2017). For example, the term *self-control* may have different meanings in different contexts (Frieze et al., 2019; Lurquin & Miyake, 2017). Equating the concept with its referent, one might conclude that there are different kinds of self-control, each constituting a measurable psychological attribute yielding a variety of psychometric procedures. Also, one may be inclined to ask whether a certain procedure actually measures self-control or something different. The latter type of question has led to an ongoing debate about the validity of psychometric procedures (e.g., Borsboom et al., 2004; Buntins et al., 2017; Newton & Shaw, 2014; Slaney, 2017).

Although critics of psychometric procedures have raised substantial issues, there is one assumption about psychological measurement that has rarely been questioned. It is the fundamental premise that psychological attributes are inherently unobservable and need to be inferred from observable behavior. Following this premise, psychometric procedures focus on the statistical relation between variables representing psychological attributes (latent variables) and variables representing observable behavior (manifest variables). The corresponding statistical methods (e.g., classical test theory, structural equation modeling, or item response theory) can be subsumed under the framework of latent variable models.

In this article, we argue that the fundamental premise of psychometrics is misleading. In our view, the main problem with psychological measurement is not unobservability; it is the fact that psychological concepts are not well-defined (see Burgos, 2021; Eronen & Romeijn, 2020). Consequently, any attempt to provide psychological measurement by solving the presumed problem of unobservability is doomed to fail unless the corresponding concept has an unambiguous meaning. We further argue that the meaning of scientific concepts is provided by the fundamental principles of a formalized theory (see Holzhauser & Eggert, 2019; Trendler, 2022). It follows that psychological measurement requires substantive formal theory. Given a well-defined concept in a formalized psychological theory, measurement consists merely in the application of the theory under standardized conditions. Therefore, substantive formal theory is both necessary and sufficient for psychological measurement. We call our approach *theory-based measurement* to emphasize that it relies on substantive theory rather than psychometric models.

The remainder of this article is organized as follows. First, we elaborate on the implications of the established modus operandi in psychometrics and on the statistical framework of latent variable modeling (LVM) in particular. The focus of our analysis is whether LVM can provide meaning to psychological concepts, without which measurement is impossible. We argue that when latent variable models are understood as mere mathematical tools (i.e., data models), the parameters in the model have no meaning over and above them being mathematical transformations of the observed data. When understood as an attempt to formalize substantive psychological theories, LVM is inherently essentialist. Consequently, the parameters in the model have no meaning over and above them being the *essence* of a certain class of behavior, and the inferred variables are nothing but empty placeholders. To illustrate our claims, we use examples from the history of science: the epicycle model for planetary motion as an instance of a physical data model, and the theory of impetus for the movement of projectiles as an example of an essentialist physical theory. We then elaborate on the relation between theory and the meaning of scientific concepts from a structuralist perspective to make our case for theory-based measurement. Finally, we outline how theory-based measurement could provide meaningful measurement procedures for psychological attributes without representing them with latent variables.

Latent variable models

According to the fundamental premise of psychometrics, psychological attributes are inherently unobservable. Therefore, the common view is that measuring psychological

attributes requires that they be inferable from observable behavior. Psychometrics offers a statistical framework that aims to justify such inference from empirical observations to unobserved psychological phenomena: LVM. The general strategy of this approach is to represent psychological attributes by a set of (random) variables called latent variables, and to represent observable behavior by another set of (random) variables called manifest variables. Latent variable models provide a mathematical description of how these two sets of variables are related. A popular way to illustrate the relation between latent variables and manifest variables is given in Figure 1. The circles represent latent variables and the squares represent manifest variables. The arrows designate the statistical relation between both types of variables, which is given in the form of a generalized linear model. This framework allows the specification of models for response probabilities, such as item response theory, as well as models for numerical test behavior, such as structural equation modeling and factor analysis.

There are various positions on how to interpret the circles and squares in a latent variable model. Some psychometricians take the position that latent variable models are nothing but convenient mathematical tools that provide a parsimonious description of psychological test data. In this view, latent variable models are data models with no further implications regarding actual psychological attributes. Data models provide an organized and, to some degree, idealized representation of data (Eronen & Romeijn, 2020). Such models can be more or less useful for practical means, but it makes little sense to say that a data model is *true* or *false*.

Others propose that latent variable models actually represent empirical relations between psychological attributes (which are taken to be unobservable) and observed

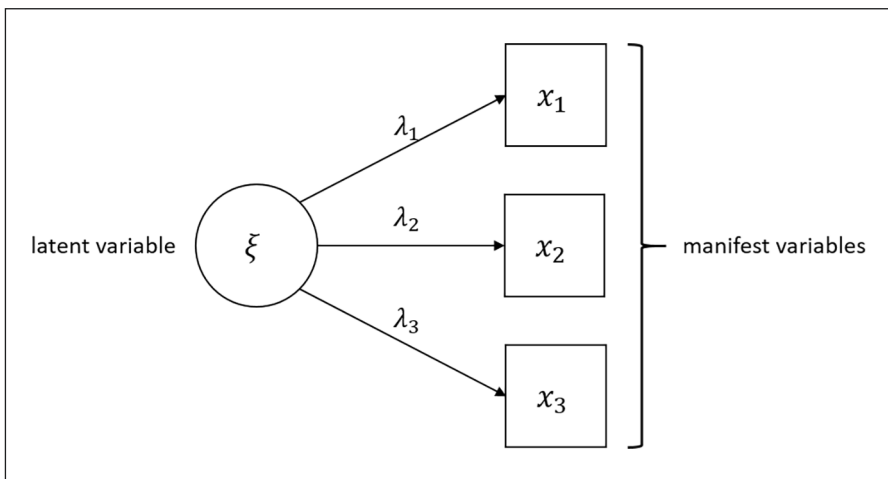


Figure 1. A simple latent variable model relating three manifest variables (indicated by the squares) to a latent variable (indicated by the circle).

Note. Latent variable models describe the relation between manifest variables and latent variables by means of a generalized linear model of the form $f(x_i) = \lambda_i \xi + \delta_i$ (the error terms are not shown in Figure 1).

test behavior. This relation is usually thought to be one of cause and effect, with the psychological attribute being the common cause of a certain class of behavior (Borsboom, 2008). In this view, latent variable models are understood as formal representations of substantive theory. Consequently, the mathematical relations specified in the model have the status of a scientific hypothesis regarding the true state of the world (Borsboom et al., 2003).

In the following sections, we will argue that neither of these interpretations can provide meaningful psychological measurements because they both imply that latent variables are meaningless. We illustrate our points by linking LVM to pre-Newtonian models of physical motion. This historical perspective shows that some of the problems encountered in modern psychology have also occurred in the development of physical theories. Therefore, it seems reasonable to explore how other sciences have dealt with these problems, and whether the strategies developed there might point to corresponding solutions in psychology.

Latent variable models as data models

From a purely pragmatist point of view, latent variable models are data models. Data models provide a convenient mathematical description of observed data, without aiming to capture the generating mechanisms behind the data. From a practical point of view, data models may be very useful when it comes to parsimonious description or statistical prediction. They are, however, not to be confused with substantive theory that aims to explain the observed phenomena (Eronen & Romeijn, 2020).

Data models have a long tradition outside the realm of psychometrics. For example, before the Copernican revolution, astronomy was dominated by the theory of epicycles. This theory describes the movement of celestial bodies by means of hierarchically nested circular orbits. An epicycle refers to a circular orbit that has its center on another circular orbit. The theory is often attributed to Apollonius of Perga (ca. 240–190 B.C.E.), although some scholars date its conceptual origins back to the Pythagoreans (Waerden, 1974). The theory of epicycles was later formalized by Ptolemy of Thebaid (ca. 100–170; Toomer, 1998). Although primarily associated with a geocentric worldview, the concept of epicycles was also used by Nicolaus Copernicus (1473–1543) when he introduced his heliocentric model of planetary movement (Dreyer, 1906/2007; Kuhn, 1971).

From an empirical point of view, the theory of epicycles works very well. It gives an excellent approximation to the observed movements of the five planets that were known at Ptolemy's time. Moreover, the elliptic orbits proposed later by Johannes Kepler (1571–1630) can be reconstructed by means of an appropriate epicycle model (see Figure 2). It was later discovered that the theory of epicycles is mathematically equivalent to a Fourier series. A Fourier series is a weighted sum of sine waves that can be used to approximate arbitrary periodic functions. Consequently, given a sufficient number of epicycles, any planetary orbit can be approximated to an arbitrary degree (Gallavotti, 2001; Hanson, 1960).

From a modern perspective, constructing hierarchically nested epicycles and adjusting their radiuses is essentially a procedure to fit an abstract mathematical model to observed data. The method of Fourier analysis can be applied to generate an approximate

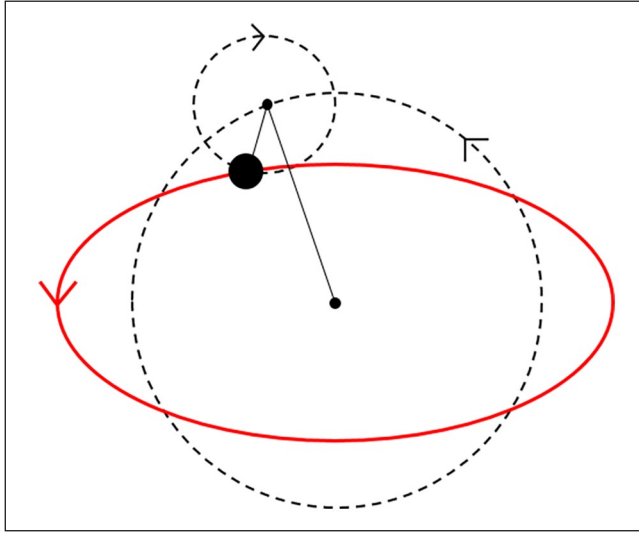


Figure 2. Epicycle model describing the elliptical movement of a planet by means of two hierarchically nested circular orbits (dashed lines).

Note. The planet moves clockwise on a small circular orbit whose center revolves counterclockwise on a larger circular orbit. The resulting observed orbit (red solid line) is identical to an elliptical orbit.

description of arbitrary planetary movements (as long as they are periodic), regardless of their underlying mechanisms. In a similar manner, we can apply the framework of LVM to generate a linear decomposition of observed random variables by means of latent variables. When understood as data models, latent variable models are best interpreted as “replacement variate generators” (Maraun, 2003). For example, the method of exploratory factor analysis will always yield a set of latent variables alongside a linear decomposition of an arbitrary set of manifest variables. Like Fourier analysis, exploratory factor analysis always provides a way to approximate the observed data to an arbitrarily close degree. The resulting model may fit the data very well, and we may use it to estimate parameters and calculate empirical predictions. However, in a certain way, the latent variables in such models are analogous to the epicycles in the Ptolemaic model. Although they may provide a convenient way to structure the data and even generate useful predictions, the parameters obtained from the Ptolemaic model have no referents. In fact, it is difficult to ascribe any theoretical meaning to the radius of an epicycle over and above the fact that (within the specified model) it provides a reasonable approximation to empirical data. The same is true for latent variables. Like the radiuses of epicycles, they have no meaning over and above being parameters in a mathematical model. Moreover, since it is always possible to account for deviations between model predictions and empirical observations by “adding epicycles” (or, in the case of LVM, “adding latent variables”), neither the theory of epicycles nor the framework of LVM is refutable on empirical grounds. Although it is possible to test specific models, the general commitment to latent variables is immune to empirical falsification (see Popper, 1935).

The analogy between latent variables and epicycles illustrates that the parameters in a data model bear no resemblance to anything outside the mathematical structure imposed by the model. Consequently, if latent variable models are data models, the estimated parameters cannot be representations of psychological attributes. They are (random) variables that are invented by the modeler to make the data more accessible or facilitate predictions about observed (random) variables. The so-constructed latent variables have no meaning over and above being descriptions of the data at hand. If they have no meaning, they cannot correspond to psychological concepts, let alone the attributes these concepts refer to. Therefore, a purely pragmatist interpretation of latent variables is inconsistent with the view that psychometrics aims to provide psychological measurement procedures. In fact, it is not even concerned with measurement in the first place.

Latent variable models as substantive theories

Although much theoretical work on LVM seems to build on the view that latent variable models are data models, many applications of LVM clearly depart from a purely pragmatist position. For example, in personality psychology, LVM is usually theoretically motivated. No one would fit an arbitrary structural equation model to a data set of testing behavior without a theoretically informed hypothesis about the underlying structure of psychological attributes. In many cases, the observed behavior as such is less interesting than the structural model one wishes to test. Therefore, in many empirical applications, latent variable models are best understood as scientific hypotheses about empirical relations between unobservable psychological attributes. In this sense, they are an attempt to provide substantive formal theories in psychology (Borsboom, 2008).

If latent variable models are understood as substantive theories, the parameters that correspond to latent variables provide numerical representations of hypothetical psychological attributes. As argued by Borsboom et al. (2003), this view implies at least some kind of scientific realism with regard to the psychological attributes one wishes to represent. Therefore, we call the position that latent variable models are substantive formal theories the *realist view* to contrast it with the *pragmatist view* that they are data models.

Although the realist view avoids some of the problems associated with the pragmatist view, it introduces a new problem: essentialism. The term *essentialism* is borrowed from Aristotle, who thought that all natural categories reflect some essential, intrinsic property. For example, an essentialist would define the category of *white things* by the shared property of *whiteness*. Similarly, the category of *tigers* would be defined by an abstract property called *tigerness*. Whereas essentialism persists in some branches of modern metaphysics, it has long been abandoned as a descriptive, let alone explanatory, mode in the natural sciences (Borgstede, 2021; Palmer & Donahoe, 1992). Essentialism is *unscientific* in that, from an empirical point of view, essences are unidentifiable. It is impossible to point to the essence of a category like whiteness or tigerness because all there is to say about whiteness is that it is shared by all white objects, and all there is to say about tigerness is that it is shared by all tigers. Therefore, if we use these essences to define the categories, the definitions become circular. If, on the other hand, we use the members of the categories to define the essences, the corresponding concepts become vacuous.

Consequently, a concept designating an essence (e.g., tigerness) is nothing but a placeholder for something unknown to the observer.

Despite these flaws, essentialist modes of explanation played an important role in premodern science. As an illustrative example, we shall take the theory of impetus. The basic idea behind this theory is that objects naturally move toward the ground unless they are moved by an externally applied force. This account worked well for objects like ploughs or carriages, which had to be pushed or pulled in order to move. However, there is no visible external force acting on projectiles like arrows while they are moving through the air. Impetus theory solved this problem by introducing a quantity that was inherent in the projectiles and moved them in a *nonnatural* direction. This quantity was called *impetus*, and was assumed to be raised by sudden external forces (like throwing or shooting) and to naturally decay over time. In this way, it was possible to *explain* the movement of projectiles as being the result of the natural tendency of objects to move toward the ground and the impetus inherent in the projectiles (Moody, 1942). However, just like whiteness or tigerness, the concept of impetus was in fact only an empty placeholder for the unknown causes of projectile movement. In other words, the theory of impetus is inherently essentialist. The category it ought to characterize is the set of moving projectiles. And the sole defining criterion of impetus is that it is an inherent property that is shared by all moving projectiles. Therefore, impetus refers to the unknown and inaccessible essence of movement in projectiles.

The theory of impetus is more than just a historical curiosity. It illustrates how essentialist thinking can find its way into scientific theories even in a down-to-earth field like ballistics, which seems less prone to metaphysical speculation than psychology. Its relevance for psychology becomes even more evident if we attempt to formalize the theory of impetus. It turns out that the theory of impetus is exactly the kind of theory that is formally represented by latent variable models.

As an illustrative example, let us consider the flight of rocks that are fired from a catapult. The catapult can be characterized by the length of the catapult arm, the height of release, the shooting angle, and the thrust of the catapult arm. The projectiles are rocks that may vary in size, weight, and shape. Every time a catapult releases a rock, we can assess the flight distance and the flight time. We can now relate these manifest variables with the latent variables implied by the theory of impetus. The flight time and flight distance are clearly indicators of impetus (*I*). The projectile mass and projectile volume can be interpreted as indicators for an object's capacity to resist external forces—that is, its resistance (*R*). And the length of the catapult arm and the height of release can be indicators of the force (*F*) that is applied when the rocks are fired. Force and resistance are hypothesized to be independent of one another and to affect the impetus of the rocks. We thus have a structural model with three latent variables and six manifest variables. Given empirical data on rocks fired by catapults, we can now test the theory of impetus by using a structural equation modeling approach.

Figure 3 shows the results of a corresponding structural equation model that was applied to a simulated data set of catapult shots using Newton's laws of motion as the data-generating model. The parameters obtained from the latent variable model are consistent with the theoretical predictions, and the overall model fit (as indicated by the *p* value and the reported fit indices) was exceptionally good.¹

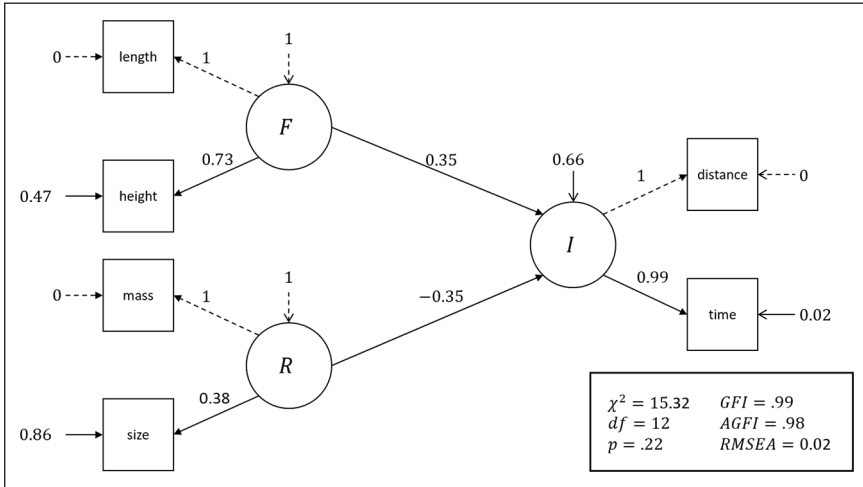


Figure 3. Structural equation model of the pre-Newtonian theory of impetus, which describes the ballistic movement of projectiles by means of three latent variables: force (F), resistance (R), and impetus (I).

Note. The model was estimated using a simulated data set with $N = 500$. The data was generated by applying Newton’s laws of motion to a simple system that consisted of catapults of varying sizes shooting rocks of varying sizes. The catapults’ arm length and release height were taken as indicators of F; the projectile mass and projectile diameter were taken as indicators of R; and the flight distance and flight time were taken as indicators of I.

AGFI, Adjusted Goodness of Fit Index; GFI, Goodnes of Fit Index; RMSEA, Root Mean Square Error of Approximation.

Imagine a world where physical theory building had stopped here, thinking that the model was a good enough approximation to the truth. The model might help to predict the impact of projectiles and possibly lead to moderate technological advances, such as the more effective construction of catapults. However, since the underlying theory is inherently essentialist, the model provides no deeper understanding about the reasons why certain catapult constructions may be more effective than others. An essentialist theory may be of some practical use but of very limited theoretical scope. Neither does the model point toward the discovery of the physical principles that produce the observed regularities. Therefore, although the model appears to advance the knowledge about ballistic movement, it is actually a scientific dead end.

The above example shows that the realist interpretation of latent variables introduces a major problem: the formal structure of LVM does not allow the inference of anything about latent variables except that they are *common factors* underlying the manifest variables. Consequently, if latent variables represent psychological attributes, all we know about these attributes is that they are some common properties underlying the observed behavior. For example, if we apply a latent variable model to a collection of verbal statements referring to aggressive behavior, we may use the manifest variables to infer a latent variable as a common factor. However, if we interpret this latent variable as a representation of a psychological attribute, this supposed attribute (if it exists) is nothing

but an abstraction of some unspecified common aspects of the manifest variables (i.e., verbal statements about aggressive behavior). We might call this hypothetical attribute *aggressiveness* and call the numerical representation obtained from the latent variable model a *measurement* of this hypothetical attribute. But what is this attribute? What does the concept aggressiveness refer to? Following the above line of reasoning, the only meaning of aggressiveness is that it refers to something unknown that underlies those verbal statements which we have categorized as instances of aggressive behavior. Therefore, if the concept refers to anything at all, it must be the essence of the behavioral category defined by the observed behavior. In other words, a hypothetical psychological attribute like aggressiveness relates to aggressive behavior exactly as impetus relates to moving projectiles and tigerness relates to tigers. It is an empty placeholder for something unknown that is inaccessible to empirical investigation.

Latent variable models fail as an attempt to provide substantive formal theories because they are inherently essentialist. If latent variables are nothing but empty placeholders, they have no meaning over and above their use as common factors in a latent variable model. Therefore, like the pragmatist view, the realist interpretation of latent variables cannot provide a sound foundation for meaningful psychological measurement.

Theory-Based measurement

So far, we have argued that psychometrics rests on the fundamental premise that psychological attributes are unobservable and have to be inferred from observable behavior. We have identified LVM as the predominant approach to provide psychological measurement procedures by attempting to solve the alleged problem of unobservability. However, latent variables are either mathematical constructions without any surplus interpretability or empty placeholders for something unknown.

Our critique of LVM is independent of the specific assumptions made in latent variable models, such as quantitative structure, ergodicity, local independence, and so forth. The main thrust of our analysis is not that the statistical framework of LVM is inadequate, but that it addresses the wrong problem.

In the following sections, we take a different approach to the problem of psychological measurement by abandoning the fundamental premise of psychometrics in favor of a semantic view of psychological measurement. From this perspective, the real problem with psychological measurement is not that psychological attributes are unobservable; it is that measurement requires observations that result from the application of substantive formal theory. Consequently, the key to psychological measurement is that psychological concepts are provided with formal theoretical embedding, such that measurement procedures can be derived from the theory.

The meaning of scientific concepts

Psychologists continue to perceive psychological attributes as unobservable counterparts to psychological concepts that are borrowed from everyday language, such as *aggressiveness*, *intelligence*, or *self-control*. The problem with this strategy is that these concepts are often ambiguous and too vague to be used in a scientific theory (Leising &

Borgstede, 2019).² For example, a common term like aggressiveness may have different meanings depending on the context. Although these different meanings may be clear enough in many contexts (i.e., people usually have no problems with using the word), it is difficult (if not impossible) to derive an unambiguous set of necessary and sufficient conditions for the correct use of the term. Although formalisms to capture the blurred boundaries of psychological concepts exist, these everyday concepts tend to invoke naive theories of human behavior as found in folk psychology, which may hinder scientific progress (Buntins et al., 2016, 2017).

Instead of adhering to the semantics associated with psychological concepts in everyday language, we propose that scientific concepts should be defined in the context of a scientific theory. Given a sufficiently developed formal theory, psychological concepts are reduced to the theoretical terms in the theory. For example, a psychological concept like *value* may be inspired by a vague common concept that we use to describe different preferences across individuals. However, when used in the context of a formalized psychological theory, the term may mean something different. For instance, in the context of reinforcement learning, the value of different behavioral options may be defined in terms of expected evolutionary fitness, rather than subjective preferences (Borgstede, 2020; Borgstede & Eggert, 2021; Borgstede & Luque, 2021). Although the theoretical term will usually overlap with the original meaning, we no longer apply the common semantics but the semantics of our theory (Trendler, 2022). In fact, the practice of deriving meaning from common language is necessarily replaced by the construction of theoretical terms when scientific concepts are sharpened according to the improvement of substantive formal theory. For example, in common language, the term *work* is only vaguely defined because it is impossible to state a definite criterion for its meaning. Instead of one general criterion that subsumes all kinds of work, there are different, context-specific meanings of work that may be more or less similar to one another. In other words, different uses of the word *work* are similar in that they bear a certain “family resemblance” (Wittgenstein, 1953/1989). In a formalized theory like classical mechanics, the term *work* is used rather differently. The meaning of physical work is simply given by the mathematical product of force and distance. There are some specific contexts where physical work coincides with the meaning of the common-language term (e.g., a farmer pushing a plough). However, the scientific meaning no longer relies on the common-language concept because its semantics are given by the formal theory of classical mechanics. In the next section, we will elaborate on how scientific theories can provide meaning for theoretical terms by focusing on the structure of scientific theories from a semantic point of view.

The structure of scientific theories

In the tradition of logical positivism, scientific theories are understood as a collection of statements about the world.³ In the case of a formal theory, these statements are to be given in a formal language (e.g., first-order logic) and are often called the *axioms* of the theory. The axioms describe the (mathematical) relation between theoretical terms (Carnap, 1995). In order to apply the theory, one has to link these theoretical terms to observational terms using unambiguous rules of correspondence. In the classical view,

these rules of correspondence provide the operational definitions for the theoretical terms of the theory in that they specify how to translate observations into the theoretical vocabulary of the theory.

This syntactic conception of scientific theories was later questioned by philosophers of science, who proposed an alternative—semantic—view (Giere, 1988; Suppe, 1989; Suppes, 1970). The semantic approach shifts the focus away from axioms and rules of correspondence toward the class of structures they represent. Logically, this corresponds to the formal semantics of the axiomatic core of a theory. For example, classical mechanics is not understood as a specific set of axioms (say, Newton's laws of motion) but as the class of structures that can be described by the fundamental principles of the theory. This implies that one and the same theory can be expressed using different sets of axioms—for example, the formalisms of Lagrange or Hamilton as alternatives to the Newtonian system.

One of the most sophisticated versions of the semantic view is metatheoretical structuralism (Balzer et al., 1987). In the structuralist view, scientific theories are described as a collection of hierarchically nested theoretical elements. These theoretical elements are related to one another in a hierarchical network—the so-called *theory net*. On top of a theory net stands the fundamental principle, which describes the most general class of structures captured by the theory. The remaining theoretical elements are specializations of this fundamental principle in that they provide additional constraining conditions under which the theory is to be applied; in other words, they narrow down the set of intended applications of the theory. According to the structuralist view, the fundamental principle is a general statement about how the theory accounts for the class of structures it is intended to explain. The various specializations of the fundamental principle provide additional information about how the theory is to be applied by describing more restrictive substructures.

The fundamental principle tells us what the theory is about, whereas the subordinate theoretical elements provide additional information about how the theory is to be applied to various empirical systems. In this sense, the fundamental principle provides a shared conceptual framework for practitioners in the field. For example, Newton's second law states that the force acting on a particle equals the product of its mass and its acceleration. Taken on its own, this law is more of a definition than an empirical fact. We cannot test Newton's second law unless we narrow down the scope of its intended applications by additional theoretical elements, such as the law of gravity, Hooke's law, Archimedes' law, and so forth. Hence, terms like *mass*, *acceleration*, and *force* are not defined with respect to anything external to the theory but by the relations specified between them within the theory. Similarly, the mathematical relation between them is not empirically discovered by applying rules of correspondence and establishing functional relations between these operationally defined terms, but is stated in the fundamental principle of the theory. Empirical tests of the theory are then constructed by using the theoretical terms and their proposed relations, and applying them to specific empirical scenarios.

Theory and measurement

In light of the semantic view of scientific theories, we realize that the meaning of a theoretical concept is not external to the theory but an integral part of the theory. The net of

all the theoretical elements that make a direct or indirect statement about a theoretical term provides its semantics.⁴ Empirical applications of the theory assign specific values to these theoretically defined concepts. The rules to obtain these values are given by those specializations of the theory that describe the structure of the intended application. For example, if we want to apply Newton's second law to describe the motion of the planets in our solar system, we need to identify the corresponding structure by additional theoretical elements, like the law of gravity. Applying the laws of motion, we can, for example, calculate the mass of a certain planet relative to the mass of other planets from their observed trajectories. Similarly, we can apply the same theoretical principles to a standardized experimental setup like a beam scale. The beam scale introduces specific constraints to the possible movements of the objects in the pans. Within these constraints (and in combination with the more specific law of the lever and the law of gravity), Newton's second law predicts that the beam scale is in an equilibrium state if and only if the objects in the pans have equal masses.⁵ Alternatively, we may attach the same objects to a spring scale in a known gravitational field (say, on the surface of the planet Earth). In this case, we can use Newton's second law to predict the displacement of the attached object relative to the displacement of other objects, given the masses of both objects. We might even use a standardized catapult, like in the example given in the second section of this article, to obtain the objects' masses. Given the constraints imposed by the specific construction of the catapult, Newton's second law predicts the flight distance of an object with a known mass (for the corresponding calculation, see Appendix 1). Therefore, we can apply the formal framework of classical mechanics to calculate the object's mass from the observed distance.

Superficially, the above procedures seem rather diverse. In the case of the beam scale and the spring scale, we are inclined to call the specific experimental setup a *measurement device* and the application of the theory to this specific setup *measurement*. However, the observation of planets or catapult projectiles is no less an instance of measurement than the act of putting objects in the pan of a beam scale or attaching them to a spring scale. In fact, every successful application of an empirical theory provides specific values for the theoretical concepts that are relevant to the application. Therefore, an operational definition is not something external to the theory but a description of a standardized experimental setup that is suitable to apply the theory. Consequently, from a semantic view, operational definitions are not definitions (in the sense of analytical statements) at all. They cannot be applied independently of a theory because, without the theory, it is impossible to connect the results of an experiment to theoretical terms (Suppe, 1972). In this sense, the above experimental procedures are all valid operationalizations of the abstract concept of mass because they relate the corresponding theoretical term to empirical observations using the formal apparatus of substantive theory that explains these observations. If the experimental setup is highly standardized and serves the sole purpose of calculating a specific value for a theoretical term, we call it a measurement device. Therefore, a beam scale, a spring scale, and even a catapult (among many other conceivable mechanical appliances) can be called measurement devices with regard to an object's mass.

Building on the structuralist conception of scientific theories, we can now give an explicit definition of theory-based measurement: theory-based measurement consists in

performing a standardized application of a theory that is suitable to provide enough information to calculate a hitherto unknown value for a theoretical term. Following this definition, an attribute is measurable if and only if there exists substantive formal theory involving the attribute, and at least one successful application of the theory that allows one to calculate a value for the corresponding theoretical term from empirical observations. The theoretical terms may, but need not, be inspired by common-language concepts. For example, the theoretical terms used in classical mechanics are mostly borrowed from common-language concepts (e.g., force, energy), whereas particle physics introduces terms that have no common-language counterparts at all (e.g., Higgs boson, charm quark).⁶ Moreover, theory-based measurement may be applied to quantitative as well as qualitative attributes, as long as they are well-defined within a substantive formal theory.

Theory-based measurement approaches the problem of ambiguously defined common-language concepts by providing an explicit semantic account of theoretical terms. At the same time, it provides a general rationale to construct measurement procedures once a theory exists. In other words, formal theory does not only provide meaning to the concepts used in a theory; it also provides rules about how to apply the theory, rendering a specific measurement theory obsolete (Humphry, 2017).

Consequently, adopting theory-based measurement, we realize that psychological measurement has proven notoriously difficult not because psychological attributes are unobservable, but because we lack substantive formal theory.⁷ In fact, all theoretical terms refer to unobservable attributes, not only the ones employed in psychology. In other words, psychology does not face a specific measurement problem, but a problem with theory (see Muthukrishna & Henrich, 2019; Smaldino, 2019). Theory-based measurement implies that once a psychological concept is well-defined within substantive formal theory, it is—in principle—measurable. Therefore, substantive formal theory is necessary and sufficient for measurement.

How to square the circle

In light of the semantic view outlined above, the primary issue in psychological measurement is that psychometric procedures provide operational definitions for psychological attributes that are themselves unidentified. We have argued that similar problems were also present in premodern physics. The theory of epicycles lacked a meaningful interpretation of the model parameters because it was a data model that merely transformed numerical observations into a new set of variables. The theory of impetus lacked a meaningful interpretation of the model parameters because it was inherently essentialist.

Both problems were solved with the development of modern physical theory—most importantly, with Newton’s laws of motion. The theory net of classical mechanics provides clear semantics for theoretical terms like *mass*, *force*, or *energy*. At the same time, the scope of the fundamental principles of classical mechanics is narrowed by various specializations of the theory, such that the theory can be applied to many diverse empirical scenarios. Some of these empirical applications have proven especially useful in determining specific values for the theoretical concepts. For example, Hooke’s law links the values obtained from a spring scale to the abstract concept of mass. Similarly, the law

of equipartition links the values obtained from a mercury thermometer to the abstract concept of energy (see Chang, 2007). Although some of these measurement procedures had been around before there was a corresponding theory, they did not yet have a scientific meaning. For example, beam scales were employed long before there was a well-defined concept of mass, mainly in the context of trade and the exchange of goods. In other words, measurement procedures had an instrumental value even though they did not have a scientific value.

As long as the utility of a measurement procedure only relies on its instrumental value, it is impossible to decide which of several procedures to measure an attribute is “the correct one.” At best, it is possible to decide which is the most useful. However, once there is an established theory, some of these measurement procedures may turn out to be superior to others with respect to their theoretical value. Moreover, new applications of the theory may provide new ways to measure the corresponding attributes. Some procedures may even turn out to be useless with regard to the theoretical concepts of the theory, and may consequently be abandoned. Theory-based measurement, as outlined above, is a conceptual framework that enables scientists to evaluate existing measurement procedures and construct new measurement procedures on theoretical, rather than statistical, grounds.

When compared to the development of measurement in physics, most psychological measurement appears to be in a prescientific stage. Although there are various procedures that produce numerical values—like psychological tests or standardized experimental paradigms—most psychological attributes have no theoretical value in the sense that they do not correspond to the theoretical terms in substantive formal theory. Given that these measurement procedures only have an instrumental value, it is not surprising that much research on psychological testing focuses on the social implications of certain test applications—for example, in the context of diagnostics, educational benchmarking, or employee selection (see Holzhauser & Eggert, 2019).

LVM can provide standards for such practical applications, but it cannot solve the problem that psychological measurement is stuck at a prescientific stage, where we apply statistical methods without sufficient knowledge of the theoretical principles underlying these procedures. Applying the framework of theory-based measurement, psychological concepts can potentially become theoretical terms in a scientific theory. Within such a theory, psychological measurement procedures could be interpreted in terms of theoretically meaningful concepts. Unfortunately, the construction of formal theory is the exception rather than the rule in psychology. Therefore, it is difficult to predict whether theory-based measurement (as it is already employed in the natural sciences) will eventually replace the current *modus operandi* in psychology. However, if we are correct in our view that the main problem of psychological measurement is not unobservability but a lack of substantive formal theory, it is indeed possible to “square the circle” by means of theory-based measurement.


Declaration of conflicting interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Matthias Borgstede  <https://orcid.org/0000-0003-1071-9038>

Frank Eggert  <https://orcid.org/0000-0002-5797-4741>

Notes

1. For details of the simulation, see Appendix 1.
2. Unless, of course, the theory is equally vague.
3. Since data models are not intended to say anything about the actual structure of the world, by “scientific theory,” we mean substantive theory in the above sense.
4. This is not to be confused with the nomological network approach put forward by Cronbach and Meehl (1955) to defend the idea of “construct validity.” Nomological networks are collections of theoretical terms that are related by scientific laws. In contrast to this view, semantic theory nets are collections of theory elements that are themselves scientific laws related by a specialization relation.
5. Notably, when applied to a collection of objects of various masses, these laws predict exactly the kind of structure that is specified by the axioms of extensive measurement from representational measurement theory (see Krantz et al., 1971).
6. It is sometimes argued that psychological measurements need to be consistent with the corresponding common-language concepts—for example, when a test procedure is evaluated against the everyday use of the corresponding term. We do not see any necessity for a correspondence between common-language concepts and scientific concepts. Therefore, following Buntins et al. (2017), we think that “validating” psychological measurement against common-language semantics is a backward step, which constrains the potential for new scientific discoveries.
7. One may argue that psychology lacks substantive formal theory because the psychological realm simply does not consist of the kind of “things” that can be measured. In our terms, this would imply that psychological theorizing can either be substantive (thereby capturing the “true” nature of the psychological realm) or formal (thereby providing means to estimate model parameters), but not both. However, such a conviction would not contradict our analysis; it would merely express skepticism toward the feasibility of theory-based measurement outside the physical realm. Although there may be good arguments for such skepticism, we prefer to be optimistic with regard to the future of psychology as a natural science.

References

- Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An architectonic for science: The structuralist program*. Reidel.
- Borgstede, M. (2020). An evolutionary model of reinforcer value. *Behavioural Processes*, 175, Article 104109. <https://doi.org/10.1016/j.beproc.2020.104109>
- Borgstede, M. (2021). Why do individuals seek information? A selectionist perspective. *Frontiers in Psychology*, 12, Article 684544. <https://doi.org/10.3389/fpsyg.2021.684544>
- Borgstede, M., & Eggert, F. (2021). The formal foundation of an evolutionary theory of reinforcement. *Behavioural Processes*, 186, Article 104370. <https://doi.org/10.1016/j.beproc.2021.104370>

- Borgstede, M., & Luque, V. J. (2021). The covariance based law of effect: A fundamental principle of behavior. *Behavior and Philosophy*, 49, 63–81. https://behavior.org/wp-content/uploads/2022/01/BP-v49-Borgstede2_corrected.pdf
- Borgstede, M., & Scholz, M. (2021). Quantitative and qualitative approaches to generalization and replication: A representationalist view. *Frontiers in Psychology*, 12, Article 605191. <https://doi.org/10.3389/fpsyg.2021.605191>
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspectives*, 6(1–2), 25–53. <https://doi.org/10.1080/15366360802035497>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219. <https://doi.org/10.1037/0033-295X.110.2.203>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Buntins, M., Buntins, K., & Eggert, F. (2016). Psychological tests from a (fuzzy-)logical point of view. *Quality & Quantity*, 50(6), 2395–2416. <https://doi.org/10.1007/s11135-015-0268-z>
- Buntins, M., Buntins, K., & Eggert, F. (2017). Clarifying the concept of validity: From measurement to everyday language. *Theory & Psychology*, 27(5), 703–710. <https://doi.org/10.1177/0959354317702256>
- Burgos, J. E. (2021). The real problem with hypothetical constructs. *Perspectives on Behavior Science*, 44(4), 683–704. <https://doi.org/10.1007/s40614-021-00311-0>
- Carnap, R. (1995). *An introduction to the philosophy of science*. Dover Publications.
- Chang, H. (2007). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Dreyer, J. L. E. (2007). *History of the planetary systems from Thales to Kepler*. Cosimo Classics. (Original work published 1906)
- Eronen, M. I., & Romeijn, J.-W. (2020). Philosophy of science and the formalization of psychological theory. *Theory & Psychology*, 30(6), 786–799. <https://doi.org/10.1177/0959354320969876>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research. *Social Psychological and Personality Science*, 8(4), 370–378. <https://doi.org/10.1177/1948550617693063>
- Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*, 23(2), 107–131. <https://doi.org/10.1177/1088868318762183>
- Gallavotti, G. (2001). Quasi periodic motions from Hipparchus to Kolmogorov. *Rendiconti Lincei: Matematica e Applicazioni*, 9(12), 125–152. http://www.bdim.eu/item?id=RLIN_2001_9_12_2
- Giere, R. (1988). *Explaining science*. University of Chicago Press.
- Hanson, N. R. (1960). The mathematical power of epicyclical astronomy. *Isis*, 51(2), 150–158. <https://doi.org/10.1086/348869>
- Holzhauser, N., & Eggert, F. (2019). The role of measurement in theorising about the world. *Social Science Information*, 58(2), 301–326. <https://doi.org/10.1177/0539018419860082>
- Humphry, S. M. (2017). Psychological measurement: Theory, paradoxes, and prototypes. *Theory & Psychology*, 27(3), 407–418. <https://doi.org/10.1177/0959354317699099>

- Krantz, D. H., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. 1. Additive and polynomial representations*. Academic Press.
- Kuhn, T. S. (1971). *The Copernican revolution: Planetary astronomy in the development of western thought*. Harvard University Press.
- Leising, D., & Borgstede, M. (2019). Hypothetical constructs. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–6). Springer International. https://doi.org/10.1007/978-3-319-28099-8_679-1
- Lurquin, J. H., & Miyake, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology*, 8, Article 568. <https://doi.org/10.3389/fpsyg.2017.00568>
- Mangalam, M., & Kelty-Stephen, D. G. (2021). Point estimates, Simpson's paradox, and nonergodicity in biological sciences. *Neuroscience and Biobehavioral Reviews*, 125, 98–107. <https://doi.org/10.1016/j.neubiorev.2021.02.017>
- Maraun, M. D. (2003). *Myths and confusions: Psychometrics and the latent variable model* [Unpublished manuscript]. <http://www.sfu.ca/~maraun/myths-and-confusions.html>
- Maraun, M. D., & Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas in Psychology*, 31(1), 32–42. <https://doi.org/10.1016/j.newideapsych.2011.02.006>
- 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: Critical history of a methodological concept*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511490040>
- Michell, J. (2009). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical and Statistical Psychology*, 62(1), 41–55. <https://doi.org/10.1348/000711007X243582>
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspectives*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
- Molenaar, P. C. M. (2008). Consequences of the ergodic theorems for classical test theory, factor analysis, and the analysis of developmental processes. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 90–104). SAGE. <https://doi.org/10.4135/9781412976589.n5>
- Moody, E. A. (1942). *John Buridan: Quaestiones super libros quattuor de caelo et mundo* [John Buridan: Questions about the four books on the heavens and the earth]. Medieval Academy of America.
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 3(3), 221–229. <https://doi.org/10.1038/s41562-018-0522-1>
- Newton, P. E., & Shaw, S. D. (Eds.). (2014). *Validity in educational and psychological assessment*. SAGE.
- Olthof, M., Hasselman, F., & Lichtwarck-Aschoff, A. (2020). Complexity in psychological self-ratings: Implications for research and practice. *BMC Medicine*, 18(1), Article 317. <https://doi.org/10.1186/s12916-020-01727-2>
- Palmer, D. C., & Donahoe, J. W. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *The American Psychologist*, 47(11), 1344–1358. <https://doi.org/10.1037/0003-066X.47.11.1344>
- Popper, K. (1935). *Logik der Forschung* [The logic of scientific discovery]. Springer.
- R Core Team. (2020). *R: A language and environment for statistical computing*. <https://www.R-project.org/>

- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Slaney, K. L. (2017). *Validating psychological constructs*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-38523-9>
- Slaney, K. L., & Garcia, D. A. (2015). Constructing psychological objects: The rhetoric of constructs. *Journal of Theoretical and Philosophical Psychology*, 35(4), 244–259. <https://doi.org/10.1037/teo0000025>
- Smaldino, P. (2019, November 6). Better methods can't make up for mediocre theory. *Nature*, 575(7781), Article 9. <https://doi.org/10.1038/d41586-019-03350-5>
- Suppe, F. (1972). Theories, their formulations, and the operational imperative. *Synthese*, 25(1–2), 129–164. <https://doi.org/10.1007/BF00485000>
- Suppe, F. (1989). *The semantic conception of theories and scientific realism*. University of Illinois Press.
- Suppes, P. (1970). *Set-theoretical structures in science*. Stanford University Press.
- Toomer, G. J. (1998). *Ptolemy's Almagest*. Princeton University Press.
- Trendler, G. (2022). Is measurement in psychology an empirical or a conceptual issue? A comment on David Franz. *Theory & Psychology*, 32(1), 164–170. <https://doi.org/10.1177/095935432111050025>
- Uher, J. (2021). Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies. *Journal of Theoretical and Philosophical Psychology*, 41(1), 58–84. <https://doi.org/10.1037/teo0000176>
- Waerden, B. L., & van der (1974). The earliest form of the epicycle theory. *Journal for the History of Astronomy*, 5(3), 175–185. <https://doi.org/10.1177/002182867400500303>
- Wittgenstein, L. (1989). *Philosophical investigations* (3rd ed.). Blackwell. (Original work published 1953)

Author biographies

Matthias Borgstede is a senior researcher at the University of Bamberg. His research focuses on the methodological foundations of measurement and the development of formal theory in psychology from an evolutionary perspective.

Frank Eggert is a professor of research methods and biopsychology at the Technische Universität Braunschweig. His primary research interests are the methodological problems of psychology and the theoretical integration of behavioral psychology and behavioral ecology.

Appendix I

Catapult simulation

We simulated a data set of 500 projectiles, with each shot from a different catapult. The catapults and rocks were randomly generated such that the catapults had an average arm length of 4.95 m ($SD = 0.48$ m) and an average release height of 5.87 m ($SD = 0.66$ m). The shooting angle and arm thrust were held constant. The rocks had an average diameter of 0.5 m ($SD = 0.01$ m) and an average weight of 66.2 kg ($SD = 10.61$ kg). To obtain the flight time (t) and flight distance (d), Newton's laws of motion were applied to the catapult system, resulting in the standard formulas for

ballistic trajectories. The equations used to calculate the flight time and flight distance for the simulated catapult data were:

$$t = \frac{v \sin(\alpha) + \sqrt{(v \sin(\alpha))^2 + 2gh}}{g}$$

and

$$d = v \cos(\alpha)t$$

respectively, with the launching angle $\alpha = 80^\circ$, acceleration due to gravity $g = 9.81 \frac{m}{s^2}$, and a launching velocity of $v = r\sqrt{2kx^2 / mr^2}$, given a catapult arm of length r that is stretched back by an average amount of $x = r / 2$, a projectile mass m , a launching height h , and a spring constant of $k = 10000 \frac{N}{m}$. The resulting average flight distance of the projectiles was 152.87 m ($SD = 65.43$ m), with an average flight time of 13.06 s ($SD = 2.74$ s). The simulated data was used to estimate a structural equation model with the above specifications using R version 4.0.3 (R Core Team, 2020) and the additional package lavaan (Rosseel, 2012). All of the variables were standardized before the analysis.

As predicted from the theory, force had a significantly positive effect on impetus ($p < .001$), while resistance had a significantly negative effect ($p < .001$). Furthermore, also in line with the theory, the flight time and flight distance were significantly predicted by impetus ($p < .001$). The measurement models for all three latent variables had an excellent fit, with all estimated paths being highly significant. The overall fit of the structural equation model was tested against the null hypothesis that the model is true using a likelihood ratio test. The resulting chi-square value was 15.32 with 12 degrees of freedom, yielding a p value of .22. Therefore, we can keep the null hypothesis of the model being the correct one. To further investigate the model fit, the goodness-of-fit index (GFI) and adjusted-goodness-of-fit index (AGFI) were calculated. Both indices yielded values close to 1 (GFI = .99; AGFI = .98), indicating an excellent fit. The root-mean-square error of approximation (RMSEA) was close to 0 (.02), indicating a nearly perfect fit of the overall model to the data.