

# A Critical Realist Critique of “Measurement in Quantitative Educational Research: An Instrumental Mistake?” by Sigve Høgheim

Leigh Price

*Inland Norway University of Applied Sciences (INN University), Norway*

## ABSTRACT

Both Sigve Høgheim and I advance a similar critique of representationalism, operationalism and construct validity as found in psychology. Our respective critiques hinge on the way that these concepts do not adequately consider the ontological question about the reality of our constructs. Høgheim suggests that we can circumvent the problem by returning to the classic definition of measurement which avoids “coding,” but I think that avoiding coding (representation, modelling, theorising) is impossible. Instead of trying to avoid it, I suggest that we embrace it, using a number of critical realist concepts such as emergence, layered reality, retroduction, judgemental rationalism and the semiotic triangle. I also argue that the current approach to operationalism reflects a deep contradiction, known as the epistemic fallacy, in which scientists reduce questions of ontology to questions of epistemology. Nevertheless, *despite* their contradictory version of science, research scientists in the fields of psychology and education still manage to discover things about the world, but in order to do so, they need to break their own rules. Høgheim wants such scientists to avoid the contradiction – the breaking of the rules of science – by remaining faithful to Euclid, Newton, and Descartes. On the contrary, I suggest that the contradiction can be avoided by changing the theory of science to a more adequate, critical realist version that better reflects the scientific practice of psychologists and educators.

**Keywords:** *critical realism; measurement in education; construct validity; operationalism; instrumentalism*

Received: May, 2023; Accepted: May, 2023; Published: August, 2023

## Introduction

The article by Sigve Høgheim is part of a long tradition of questioning the scientific status of the so-called “soft” or “intermediate” sciences which often relate, in one way or another, to the study of people, such as the psychological sciences, anthropology,

---

Correspondance: Leigh Price, e-post: leigh.price@inn.no

© 2023 Leigh Price. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

*Citation:* Price, L. (2023). A Critical Realist Critique of “Measurement in Quantitative Educational Research: An Instrumental Mistake?” by Sigve Høgheim. *Nordisk tidsskrift for pedagogikk og kritikk, Special Issue: Måling i kvantitativ utdanningsforskning*, 9, 174–194. <http://doi.org/10.23865/ntpk.v9.5660>

critical theory, the social sciences including education science, and the humanities generally. However, not all the disciplinary fields that some consider to be non-scientific are related to the study of people; disciplines such as climate science and ecology also fall into this category. According to Roy Bhaskar (1989), the general criterion for inclusion in this category is not that a discipline deals with people, but rather that it studies “the confluence of two or more orders of determination” (p. 91). When Bhaskar talks of these disciplines as having “orders of determination,” he is referring to his concept of “emergence,” in which lower order *entities*, such as individual neurons, individual people or individual living organisms interact synchronously in such a way as to form *emergent entities* that are not reducible to the individual entities that constitute them. Therefore, psychologists study the mind, which is emergent from neurons; social scientists study society and its structures, which are emergent from people; and ecologists study ecosystems, which are emergent from organisms. To this list we must also add climate scientists, who study climate, which is emergent from the activity of atmospheric molecules, the characteristics of which we measure daily as weather. The reason intermediate sciences are not considered to be “proper science” is because the emergent component of what they are interested in is not directly measurable or observable. That is, we cannot measure social structure, we can only measure people; we cannot measure ecosystems, we can only measure the organisms that constitute them; we cannot measure climate, we can only measure current weather characteristics such as temperature and historical records of such. The number of times that the intermediate sciences have been accused of not being proper science is too great to list comprehensively, but one gets the general gist of the accusation from, for example, Karl Popper (1961), who criticised the theories of Sigmund Freud, representing psychology, as “soothsaying” (p. 37), and the theories of Karl Marx, representing critical theory, as “pseudoscience” (p. 37–38).

I have suggested elsewhere that the argument that climate science is not proper science has seriously undermined efforts to address the climate crisis (Price, 2019). In this paper, I will make a related suggestion, which is that the argument that contemporary measurement in educational science is not proper science – represented here by Høgheim – will seriously undermine the ability of educators to provide good quality educational interventions. I will therefore use Høgheim’s article as a useful illustrative example of that which I am mainly interested in critiquing, namely scepticism about the scientific credentials of the intermediate sciences generally, and education science more specifically.

A key component of the argument in Høgheim’s article, which owes a significant debt to the work of Joel Michell (see, for example, Michell, 1993, 1997), is that education research needs to return to a classical version of measurement as outlined by Euclid, Newton, and Descartes. This argument is made in opposition to the modern or representationalist version of measurement, as originally outlined by Norman Campbell, where representation is by the codes mentioned by Høgheim (1920, as cited in Høgheim, 2023). Høgheim suggests that, in terms of this dichotomy between

measurement as coding and measurement as quantification, measurement as quantification is preferable, with coding being questionable because it is not about anything, that is, it is not an example of ontological realism. He states:

Based on a comparison with a classical understanding of measurement, it is argued that quantitative educational research does not measure, but encodes theoretical concepts. This understanding of measurement casts doubt on other prevailing beliefs in quantitative research, especially ontological and scientific assumptions. Based on the review of psychological measurement, it is argued that educational research has no connection to ontological realism. (Høgheim, 2023, my translation<sup>1</sup>)

Høgheim reflects the prediction made by Bhaskar, that “empiricists are [...] impaled on the dilemma of abandoning either their phenomena or their analysis [...] She or he can either opt for the position that nothing governs phenomena in these cases, so that nature becomes radically and capriciously indeterministic (weak actualism<sup>2</sup>), or elect that science has as yet discovered no laws—the heroic ‘strong actualist’ line” (Bhaskar, 1986/2009, p. 18). Høgheim opts for the *heroic strong actualist* line when he states, “relationships under investigation (by psychologists and educators) are based on theoretical laws which have not been discovered.” Høgheim therefore takes the Popperian position that social science (and hence education) should not emulate the natural sciences because the objects of its investigation are not discoverable, that is, not amenable to “scientific” experimentation and measurement.

I would argue that psychological measurement represents an instrumental mistake in emulating the method of the natural sciences without the methodological principles, which are neither useful nor scientific. Measurement becomes an instrumental task and the error refers to a practice that goes against its own underlying assumptions with a reduced potential for discovery. (Høgheim, 2023)

Since “discovery” is the aim of science, and since Høgheim thinks here that education science has a reduced potential for discovery, he is suggesting that educational science is not, in fact, a proper science. The position that social sciences such as education are not real science is often justified by reference to Karl Popper’s (1959/1961) demarcation criteria – demarcating the line between science and pseudoscience. When Høgheim talks about the “scientific task,” I assume that he is referring essentially to Popper’s demarcation criteria of science, which is that if something is not measurable, it is not proper science. To the contrary, Bhaskar (1979/2014) argues that we can use the same philosophy of science – but not the same methods – to

---

<sup>1</sup> All quotes by Høgheim (2023) are my translation from the original Norwegian.

<sup>2</sup> Bhaskar (2016, p. 24) defines actualism as “The collapse of the real to the actual.” The real includes the structures and mechanisms of reality, which we use theories to define; and the actual includes events, often in the form of constant conjunctions, which are typically presented as statistical correlations. We could therefore, in a narrower sense, define actualism in research as the collapse of the structures and mechanisms of the world to statistical correlations.

guide all science whether natural or social; that is, his approach to social science (and hence educational research) is a naturalistic one and there is no strong demarcation between the so-called natural sciences and the social sciences. To do this he had to change the philosophy of science so that it included the nonempirical layers of reality which we can only know about through retrodution. “A retroductive argument asks what would, if it were real, bring about, produce, cause or explain a phenomenon.” (Bhaskar, 2016, p. 3). Bhaskar therefore insisted that retrodution, in addition to deduction and induction, should be part of the scientific canon.

Bhaskar’s transcendental critical realism therefore differs from Høgheim’s argument in that he (1993/2008, p. 323) takes a stand against “substantive [...] Newtonian-Euclidean-Aristotelian [...] epistemological commitments” and therefore he is against the necessity for knowledge to depend on *apriori* truisms, such as the truth of Euclidean geometry, or Høgheim’s truisms about measurement. Similarly, Bhaskar – unlike Høgheim – is committed to coding, where coding involves categorising/making general concepts/models and the codes are assumed to refer to real entities, so that another way of describing critical realism is to say that it is committed to, amongst other things, categorical realism. According to Bhaskar (2016, p. 125), “Categorical realism is important because philosophers, especially from Kant onwards have regarded the categories as things we impose on the world, subjective impositions on being rather than inherent in being itself.” Instead, Bhaskar suggests that categories are not merely human inventions (although there is a human element to the discovery of knowledge), but rather that categories are about reality, even when the reality in question is, perhaps merely temporarily, unobservable.

Therefore, in this paper I will explore the ideas presented by Høgheim through the perspective of critical realism. I begin by situating Høgheim’s paper in terms of its contribution to various related contemporary debates on the scientific status of the intermediate sciences, which can be traced to discussions around operationalism, or the defining of measurement as a set of methodological operations. Operationalism is, one might say, agnostic about reality and thus has no ontology for what it is measuring (Nordahl-Hansen & Kvernbekk, 2020, p. 92). Like Høgheim, I will show that the application of operationalism in psychological and educational science reflects a deep contradiction. However, I will frame the contradiction differently from Høgheim and describe it as an example of the epistemic fallacy, in which questions of ontology are reduced to questions of epistemology (Bhaskar, 2016, p. 6, 23), also described as a situation in which it is assumed that “statements about being can always be analysed in terms of or reduced to statements about knowledge” (Bhaskar 2016, p. 23).

I will argue that we can avoid the epistemic fallacy by using the critical realist ontology and the concept of retrodution to creatively theorise about real, emergent entities such as human beings, ecosystems and societies, and by using judgemental rationalism to consider competing theories, choosing to run with the theory that best fits most/all of the evidence. In order to make this argument, I will provide the reader with some concepts taken from critical realism, such as: the role played by closed

systems in empiricist<sup>3</sup> science; categorial and causal realism; the idea that reality is layered with higher order levels of reality emergent from lower-level orders; and the idea of the semiotic triangle. I will also present a discussion of instrumentalism and how a critical realist approach allows us to avoid it.

### **Situating Høgheim’s paper in the context of the debates**

These debates, about the scientific credentials of the intermediate sciences, are centred on all or some of the key characteristics of Popperian so-called post-positivism, which is a variation on the theme of positivism and is therefore a kind of Humean empiricism (Bhaskar, 1986/2009, p. 80). Recently, the argument has gained attention in what has come to be known as the “crisis of replication.” This “crisis” has emerged in light of the discovery that, contrary to what one would expect according to Humean science, a negligible amount of scientific research has been replicated in, for example: psychology (Wiggins & Christopherson, 2019); the social sciences and education (Wiliam, 2022); and ecology (Filazzola & Cahill, 2021).

An earlier version of the debate, and one to which Høgheim’s paper directly refers, can be traced to the report by the Ferguson Committee (Ferguson et al., 1940), which concluded that psychological attributes cannot be measured scientifically. An important member of the Ferguson Committee was Norman Campbell, whose ideas are central to Høgheim’s paper, either directly (Høgheim refers to Campbell & Fiske, 1959; Campbell, 1920; Shadish, Cook & Campbell, 2002) or indirectly, via the work of Michell (Høgheim’s reference list includes nine papers by Michell). Michell, influenced strongly by Campbell amongst others, argues that “psychometrics (as it is currently, typically taught) actually subverts the scientific method” (2001, p. 211). He says this is because “The attributes that psychologists aspire to measure are not, as yet, directly observable. Psychologists may observe quantitative effects (e.g., test scores or reaction times) of such hypothesised attributes, but the attributes themselves are hidden from view.” He goes on to say that if one combines the lack of direct observability with the fact that psychology cannot be conducted experimentally in closed-systems (that is, in laboratories) to control for factors confounding the problem, one faces the problem that “in psychometrics, the information gleaned from quantitative effects is ambiguous” (p. 213). The novelty of the paper by Høgheim is that he shifts the arguments presented by Michell, originally levelled at the field of psychology, to the field of education. This shift is possible because both fields are intermediate sciences.

A related debate which also forms part of the historical backdrop to Høgheim’s paper is the debate in psychology around the concept of *operationalism*, exemplified

---

<sup>3</sup> I distinguish between “empirical” science, which I have no problem with, since all science must have some basis in the empirical, and “empiricist” science, which problematically reduces science to the study of empirical phenomena alone.

by the 1945 special issue of the *Psychological Review* dedicated to the discussion. A key advocate of operationalism was the behaviourist B. S. Skinner, well-known to educators. This is how the idea of operationalism in psychology is described by Bridgman, who introduced it:

We evidently know what we mean by length if we can tell what the length of any and every object is, and for the physicist nothing more is required. To find the length of an object, we have to perform certain physical operations. The concept of length is therefore fixed when the operations by which length is measured are fixed: that is, the concept of length involves as much as and nothing more than the set of operations by which length is determined. In general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations. (Bridgman, 1927, p. 5)

Here Bridgman’s operationalism reflects the epistemic fallacy in which questions of ontology (the actual existence of length) are displaced by, one could say reduced to, epistemology (operations). Nordahl-Hansen and Kvernbekk (2020, p. 91) also give an example of the epistemic fallacy when they say, “The meaning of a claim is identical to its method of verification.”

However, operationalism has, in this debate, been severely criticised along much the same lines that Høgheim has criticised it. The first-generation operationalists have therefore given way to the second generation of operationalists, amongst whom I think Høgheim can be counted as a member. In first generation operationalism, the question of whether a measurement method is valid is moot, since, if the method of measurement completely defines the concept, then there can be nothing more to it. For example, if passing an exam about computers is what is meant by knowing about computers, then there is nothing more to it and the method of measurement is valid, or as Nordahl-Hansen and Kvernbekk (2020, p. 91) describe it, “The strong version of the verification principle demands the complete translation of theoretical terms T into observational terms O (a perfect operational definition,  $T = O$ ), something which, if it were possible, would render the T-term a complete summary of the observations.”

Measurement therefore only becomes a problem if the concept, for example, knowledge of computers, is assumed to mean more than what the computer test is directly testing which, of course, it usually does. It is this question that has been directed as a criticism against first generation construct validity, and that has led to the formation of the second-generation of operationalists, who generally offer at least two solutions to it, both of which Høgheim mentions.

The first solution is recognisably coherentist, as outlined philosophically by, for example, Willard Van Orman Quine (1908–2000) and more recently, Nicholas Rescher (1928–); the second is recognisably pragmatic, as outlined philosophically by, for example, William James (1842–1910), who emphasises validity as use-value and Richard Rorty (1931–2007), who emphasises validity as consensus. Note here



that I do not mention the pragmatist Charles Sanders Peirce, since there are different kinds of pragmatisms, and which pragmatism one subscribes to is significant.

In terms of the debate in psychology, an example of the coherentist approach applied to construct validity is provided by Cronbach and Meehl (1955), who argue for a kind of network of indications that the construct is valid and use the Binet test as an example, stating that it would not have been considered a good instrument to measure IQ if its results had not matched the opinions, or cohered with the opinions, of teachers (p. 286). An example of the pragmatic approach to the problem of construct validity is given by Høgheim himself, as part of his resolution of the problem, when he offers a pragmatic defence of first-generation operationalism by arguing that even though these constructs are not about anything, they are, nevertheless, “practically applicable,” by which he means, I assume, useful. Messick (1998, p. 36) also provides an example of the pragmatic version as follows, and please note that he assumes, as does Høgheim, that many of our important constructs have no referent, that is, they are simply “heuristics” with no basis in reality as such. He says:

Just as on the realist side there may be traits operative in behavior for which no construct has yet been formulated, on the constructive side there are useful constructs having no counterpart in reality. These latter instrumental constructs are usually inductive summaries of data that serve as heuristic devices for organizing observed relationships with no necessary presumption of real entities underlying them. Examples include higher-order constructs such as “ego” or “self”, as well as useful classifications such as “working class” and “middle class” or “childhood” and “adolescence.” (Messick, 1998, p. 36)

This progression, or evolution, of the question of validity in the intermediate science of psychology, from first generation to second-generation has also been noted by Borsboom et al. (2004, p. 1061; see also Nordahl-Hansen & Kvernbekk, 2020). Another characteristic of some of the second-generation operationalists – perhaps differentiating them enough to call them third generation – is that they are what Bhaskar would call anti-deductivists, who also, like Bhaskar, are committed to the formation of theory by the logic of retroduction. A well-known anti-deductivist in the broader philosophical community is Rom Harré (1927–2019) (Bhaskar, 2016, p. 37).

Operationalists who follow this anti-deductivist perspective have something in common with another kind of pragmatism, specifically the pragmatism of Peirce, since it was he who popularised the concept of retroduction, originally taken from Aristotle (Hanson, 1958, p. 85). For instance, Borsboom et al. (2004, p. 1062) argue that correlations between test scores and other measures merely provide circumstantial evidence for certain attributes and that what needs to be tested is not a shallow theory about the relation between the attribute measured and other attributes but a deep theory of response behaviour. This, they say, is because the causal role of the attribute (which I assume we would know about by retroductively derived theory)

links together, or occurs between, one relevant factor and another. They say, “It is important to note that [...] the problem of validity cannot be solved by psychometric techniques or models alone. On the contrary, it must be addressed by substantive theory. Validity is the one problem in testing that psychology cannot contract out to methodology” (p. 1062). By refusing to “contract out” validity to methodology by insisting on theory, Borsboom et al. are, to an extent, refusing the epistemic fallacy which, as we have seen, reduces questions of ontology to questions of epistemology. As such we can call Borsboom et al. “anti-deductivists” because of their refusal to accept that validity reduces to testable *deductive* methodological questions; it also needs retroductive theory, which is technically not *deductively* testable (we can only deductively test hypotheses, not theories). Høgheim recognises a similarity here between Borsboom’s anti-deductivism and critical realism; and he is correct to do so. The anti-deductivist Harré was Bhaskar’s PhD supervisor, and Bhaskar (2016, p. 21) acknowledges that anti-deductivism is possibly the closest antecedent to his transcendental realism.

Nevertheless, Borsboom et al.’s commitment to retrodution and theory does not mean that they manage to fully avoid certain aspects of positivism, such as its actualist insistence on validity being suggested by correlations: “We have proposed a simple conception of validity that concerns the question of whether the attribute to be measured produces variations in the measurement outcomes” (Borsboom et al., 2004, p. 1069). That is, for Borsboom et al., changes in the attribute being measured are correlated with changes in the measurement outcomes. Thus, like other anti-deductivists such as Harré (Bhaskar, 2016, p. 37), they go so far as to understand that correlations are not sufficient to understand questions of causation and validity – they assume that we also need retroductive theory – but they still think that correlations are necessary. On the contrary, for Bhaskar (2016, p. 37) correlations are neither necessary nor sufficient to validate a theory of causation, whether in the natural or social sciences, which is fortunate given that experimental correlations need closed systems to occur and yet much of what interests us cannot be placed within closed systems.

### **Introducing transcendence into the debate**

The link between the different approaches mentioned so far is the question of transcendence. This question was central to the development of critical realism, so much so that it was originally called *transcendental realism* by Bhaskar. Both first and second generation approaches to operationalism and construct validity address the problem of transcendence, specifically the question of whether or not transcendent things are real. Generally, both first and second-generation operationalists are, if not irrealist about transcendent things, at least agnostic about them (Nordahl-Hansen & Kvernbekk, 2020, p. 92). For instance, Michell’s (2001 p. 211) argument that we cannot test unobservable, that is transcendent, entities, which includes emergent



entities, in laboratory-stye experiments, suggests that he is, at the very least, agnostic about the reality of transcendent things. Transcendent things can be, amongst other things, generalities/codes/universals, also known philosophically as *categories*, such as a general idea of the trait of diligence (which we recognise by characteristics such as always finishing tasks); or *causes*, such as that peptic ulcers are caused by *Helicobacter pylori*, or that stress causes students to perform poorly in examinations. We are best able to test the validity of our theories about the existence of unobservables in laboratories. This is because laboratories allow us to carry out experiments in closed systems. Why are closed systems important to achieve relative certainty about unobservables? I will answer this question first in terms of categories and secondly in terms of causes. This discussion about *categories* and *causes* also provides a conceptual backdrop to my discussion about instrumentalism.

### Categorial realism

Starting first with *categories*, let us compare the measurements of a simple natural science trait relating to heat energy (categorised as temperature) with an intermediate, non-simple, human science trait relating to the ability to reason (categorised as intelligence), neither of which we can decide accurately without the help of instruments. In terms of the atmospheric category of temperature, we usually decide temperature by looking at the movement of red-dyed alcohol or mercury in an instrument called a thermometer. How can we calibrate a thermometer to be certain that X movement in the mercury represents exactly Y change in temperature? Convention states that we use the boiling and freezing points of water at sea level. The reason that sea level is specified is because pressure, which changes with altitude, affects the result, but even sea level experiences changes in pressure, so simply carrying out the calibration at sea level is problematic if one wants a precision calibration. Therefore, to achieve precision in our temperature-gauging instrument (thermometer), we need to calibrate it in a laboratory where all the confounding factors, but especially pressure, can be controlled.

Now, if we consider the psychological category of intelligence, we often decide intelligence by looking at the results of an instrument such as the Stanford-Binet Intelligence Scale. As with temperature, there are confounding factors that can influence the manifestation of the trait of intelligence, such as socioeconomic status (Chapman et al., 2014). However, unlike temperature, we cannot use laboratories to control confounding factors in our testing of the construct validity of our instruments to measure intelligence. This is true of all instruments used for measurement in the intermediate sciences where the entity being measured is emergent from something else. In this case, intelligence is an emergent entity, a subset of mind, which is emergent from both body and, to a certain degree, I would argue, society. For ethical and logistical reasons, we cannot include these entities in laboratory experiments to test the construct validity of our intelligence tests. When researchers assume that their instruments are measuring what they think they are measuring, without

first checking “construct validity,” they are making the “instrumental mistake” that Høgheim correctly criticises.

### Causal realism

Looking secondly at the oft-times unobservable *causation*, the problem is not unlike the problem that we have with generalities, in that if the unobservable causation at hand is a non-emergent entity that exists at a scale able to be placed in a laboratory, then empirical natural science can reliably use correlations as evidence of cause and effect because confounding factors can be managed in the closed systems that laboratories allow us to create. However, if the unobservable is an emergent entity, such as intelligence or the ability to pass a mathematics examination, such management of confounding factors is difficult, if not impossible, to achieve. Therefore, any correlations that are discernable *in situ* must be interpreted in terms of an open-system context where there are many confounding factors and where a correlation between two things does not necessarily mean that there is a causal connection between them. When correlations in open systems are considered to be causal, simply because they exist, that is, when the reason behind the correlation is not considered, this is a kind of instrumentalism because no effort is made to be certain that the correlation is measuring what it is assumed to be measuring. Therefore, it might be assumed instrumentally that the correlation that exists between drinking red wine and longevity means that red wine causes people to live longer, but no attempt is made to ascertain why, or even if, this is actually the case and it may simply be that red wine is the drink of choice of wealthy people who also happen to be able to afford better quality food and better health care.

### **Categorial and causal realism are related**

Of these two kinds of realism, Høgheim’s argument is mainly relevant to categorial realism, but the argument that I make here is relevant to both types, that is, both categorial and causal realism. Indeed, identification of causation is often the precursor to the categorial naming of things. For example, in the case of water, the understanding of what causes water, namely the joining together of hydrogen and oxygen to create a molecule symbolised as  $H_2O$ , has become a way of categorising the entity also designated with the word *water*; thus, what causes water ( $H_2O$ ) has become a way of naming the category also known as water. Another example is the way that we categorise the disease often known as AIDS (acquired immune deficiency syndrome). The early designation of “AIDS” referred to certain disease symptoms, but the disease is now often referred to as “HIV,” that is, in terms of what causes it (human immunodeficiency virus).

Acknowledging that deep reality consists of both real categories and real causes can provide the conceptual tools to avoid a common error in research about causation which assumes, from the perspective of experimental science, that the stronger the

correlation, the more likely the two factors under examination are causally related. Whilst it is usual for statisticians to acknowledge that the direction of causation may be difficult to determine (for example, do high levels of stress cause poor grades or do poor grades cause high levels of stress?), they do not often discuss the possibility that a strong correlation between two factors may simply be due to the two factors being different empirical manifestations of the same phenomenon. How could they if they do not have an ontology for the phenomenon, which is an unobserved thing? For instance, school absenteeism and poor grades are strongly correlated (Morris & Rutt, 2004). One could ask: does absenteeism cause the poor grades or do poor grades *cause* absenteeism? However, equally, one could ask: are poor grades and absenteeism empirical manifestations of the same thing, namely the structural, emergent, transcendent *category* poverty? If the latter is the case, then their correlation could be due to the factors absenteeism and poor grades being empirical manifestations of the same category (the category known as poverty), rather than being causal of each other. If the entity that underlies both poor grades and absenteeism is poverty, then it makes sense to rename the offending issue “poverty.” Of course, ideologically this challenges the status quo since it shifts the blame from supposedly irresponsible parents and lazy children to the structural causes of inequality and poverty. This is, therefore, an example of how the current irrealist ontology suits those who have vested interests in maintaining the status quo.

## Retroduction

The key logic used in both categorisation and identification of causation is *retroduction*. For example, cognitively, if the image being sent to my brain by the person before me looks like my friend Aksel, I will assume it is Aksel (but I may be wrong, it may be Aksel’s identical twin Kurt). Nevertheless, I could also say that the presence of my friend Aksel in front of me is causing my eyes to send an image of Aksel to my brain and there is usually a constant conjunction of events (in this case, a causal correlation) between, epistemologically, seeing the image of Aksel and, ontologically, the presence of the real human Aksel.

Note that cognition is therefore also about causation, hence, again, the difference between these two considerations – cognition and causation – is not as great as one might initially think, and both are able to be understood in terms of retroduction. Michell (2001, p. 211) is aware of this when he mentions that not being able to measure things directly is not so much of a problem for the natural sciences because they are able to measure things indirectly in closed systems in laboratories. We can perhaps better understand this if we consider the example of the discovery of gravity. It was only when gravity was measured in isolation from the counteracting forces of air, that is, in a closed system, that the relationship between mass and gravity could be properly understood. In other words, the regular effect of gravity in the absence of confounding factors, that it always draws objects towards the earth in a certain

way (there is a constant conjunction of events or a correlation, between untethering an object above the earth and its falling to the ground) that allowed us to retroductively surmise the existence and nature of gravity itself, without directly measuring it. It is the fact that we cannot place the subject matter of the intermediate sciences into a closed system and consider it in terms of causation that makes it impossible to measure in the classic way. To put it another way, in open systems, the measurement of unobservables such as categories or causation by use of a proxy measurement has challenges which, in my opinion, have not been satisfactorily addressed by second-generation operationalists.

### **What are the implications for my debate with Høgheim?**

We can say that both Bhaskar and Høgheim question Campbell’s instrumentalist version of neo-Kantian, irrealist representationalism – which is not unlike Bridgman’s operationalism<sup>4</sup> – where the basic principle is that a term is synonymous with the way it is identified (Bridgman, 1927, quoted by Høgheim 2023; Campbell, 1928; McGrane, 2015; Bhaskar, 2016, p. 46). Thus, one has a term and a way of identifying it, but the real object being represented is missing or, as stated by Høgheim (2023), there is nothing “inherent in the reality that can be discovered.” However, Bhaskar and Høgheim offer different solutions to this problem. I will consider Høgheim’s solution first and offer a critique of it. Then I will follow this critique with Bhaskar’s solution.

Høgheim’s solution to the problem that there is no ontology underlying the theories and codes in psychology and education is, as already mentioned, to encourage education research to return to classical measurement and, thus, avoid any kind of social representation. There are at least two ways in which his position contains certain contradictions, in terms of being: (a) against representationalism but using representations; and (b) questioning the reality of the objects of measurement in theory, but not (one assumes) in practice.

#### **(a) Against representationalism but using representations**

In this article, Høgheim argues, from the position of empiricism, that psychologists are committing the “error” of representationalism; however, this is a contradiction because Høgheim and indeed all empiricist scientists must necessarily also commit that same “error.” For instance, it is assumed that “the scientific task of measurement [is] superfluous [when] it is the researcher who defines what is measured and how

---

<sup>4</sup> Campbell’s and Bridgman’s perspectives, which in one way or another reduce ontology to epistemology, were brought together for psychologists, and solidified into the concept of “operationalism” (a word coined by Bridgman but latter denounced), by the work of Stanley Smith Stevens (McGrane, 2015).

it should be measured” (Høgheim, 2023). However, “proper” measurement is then described by Høgheim (2023) in this way: “When measuring a person’s length (X) we can use a *standard* meter such as level (Y; unit) to detect the relationship between length of the person and r number of meters” (my emphasis). Yet, surely, it must be acknowledged that what is seen as proper measurement by the author is a representation by *standard* things, and this standard representation must have at some time been defined by humans. For instance, the *standard* unit used to represent (measure) length in Western societies has changed over time and is not the same as in non-Western cultures. Think here about how the Inca’s standard definition of distance was the *ricra* instead of the metre (based on the distance from fingertip to fingertip of an adult person’s outstretched arms, about 1.6 metres), and the standard system of measurement used by the Aztecs and Mayans was the vigesimal system of measurement, with 20 instead of 10 as its base (Hamilton, 2018; Stodola, 1971). It seems to me that we cannot avoid some kind of socially defined representation, no matter what we are categorising or counting (Wittgenstein, 1953/2010). Bhaskar (2016, p. 24) calls this the transitive dimension of reality, that is, the aspect of reality that is known about because of human and social ways of seeing things. He, however, acknowledges that, for the most part, human beings assume, as does Høgheim, that there is no difference between what we measure and what exists and he calls this the “natural attitude [...] in which we do not distinguish ontology and epistemology, but merely talk (in an undifferentiated way) about the known world, a standpoint that Hume and Kant merely reflected” (Bhaskar, 2016, pp. 6–7). It seems that Høgheim is suggesting a return to this “natural attitude.” Bhaskar explains that this natural attitude is not able to be maintained during times when science is in revolution and there are competing claims about what “is” (Bhaskar, 2016, pp. 6–7). However, it is at these times that researchers are active, since all research, one assumes, is aimed at developing and thus actively changing or revolutionising current knowledge.

**(b) Questioning the reality of objects of measurement in theory, but not (one assumes) in practice**

One of the problems with the natural attitude, as reflected in the classic approach to measurement followed by Høgheim, is that it is based on the epistemic fallacy which, you will recall, conflates epistemology and ontology (Bhaskar 2016, p. 31). Høgheim’s position can be seen to reflect the epistemic fallacy in this statement:

What is also worth noting is that realism in this context refers to the existence of quantitative *attributes*, not “quantitative” objects. If a person is measured, it assumes not an objective existence of “person”, but universal attributes of the person (Michell, 1999), such as height, mass and temperature. (Høgheim, 2023)

Here we see that Høgheim is, ironically, being irrealist, since for him, only the subjective, epistemological, universal attributes of a woman (her height, mass and

temperature) can be assumed to exist, and not the objective tall or short, heavy or light, warm or cold person herself. He therefore reduces the ontological (the person) to the epistemological, or, in other words, he reduces statements about being (about the person) to statements about knowledge (about height, mass and temperature). That is, he is realist about what the attributes measure, but not realist about the object being measured.

This epistemic fallacy is the standard approach taken by scientists in their empiricist theories of science, but in the practice of science these scientists must act as if the person actually exists, apart from their measurement of her. Therefore, in their practice, they act as if being (ontology) is different from epistemology (knowledge about being); that is, in their practice, they must act as if there *is* a difference between: (a) the *representation* or coding (by words or numbers) arrived at by the epistemological, knowing subject; and (b) the ontological thing itself, despite their claims otherwise. Bhaskar describes the tendency towards this contradiction thus:

Bachelard remarked on the striking *décalage* or discrepancy between the diurnal philosophy of scientists, that is the philosophy implicit in their spontaneous practice, and the nocturnal philosophy of philosophers. But what is more striking is that it is to the nocturnal philosophy of the philosophers that scientists tend to return when they self-consciously reflect upon their conscious practice. Newton, Engels, Freud, Einstein in different ways attest this phenomena. What explains the discrepancy? How are we to account for the fact that even, and sometimes especially, the greatest scientists seem systematically deluded about the nature of their work? (Bhaskar, 1986/2009, p. 152)

### **Bhaskar’s solution to the problem**

Bhaskar resolves the “problem” of representationalism by assuming categorial and causal realism. He thus argues that our theories are about something real, that is, our codes, categories and deep causal explanations are about something, and that something may be observable or unobservable, or perhaps it is merely unobservable now, but may be observable later, with the advent of new technology.

But critical realism breaks with Campbell’s neo-Kantianism by allowing that, under some conditions, these concepts or models could describe newly identified deeper, subtler or otherwise more recondite levels of reality. Theoretical entities and processes, initially imaginatively posited as plausible explanations of observed phenomena, could come to be established as real through the construction of sense-extending equipment or of instruments capable of detecting the effects of the phenomena. (Bhaskar, 2016, p. 46)

More fully, Bhaskar’s (2016, p. 34) solution is to adjust the neo-Kantian theory of representation, by including a real object that is being represented. The result is the semiotic triangle. Therefore, the critical realist semiotic triangle includes: (a) the object being represented; (b) the word/code/representation for the object;



and (c) the rules or picture in the knower's mind. The semiotic triangle is essentially the same as the approach to representation found in psychology/education which Høgheim is criticising, that is, "the rule-based assignment of numbers to objects or events," although the semiotic version is broader than this definition, as it is the rule-based assignment of numbers and/or words to objects or events ("coding" is involved, whether or not that which we know about can also be counted and thus "measured"). Explicit in the semiotic triangle is the understanding that these coding "rules" exist in the knower's/knowers' mind/s. Therefore, for example, we have rules as to what words we assign to biological entities with leaves: if they have lignified structures, they are called "shrubs" or "trees," but if they lack lignin in their structural components they are called "herbs." Some less biologically minded people will simply think of herbs as being smaller than trees and shrubs (they have a different set of rules or pictures in their heads).

### **Empiricist science needs "blind spots" to appear workable**

Bhaskar (2016, p. 38) further explains that it is because scientists do not do what they tell others they do that enables their mistaken theory to appear to work in practice. In other words, *despite* their incomplete theory of science, scientists still manage to be successful in finding out things about the world. As would be expected if the critical realist view is correct and empirical realism is problematic, any version of applied research, based on empirical realism, that does not break its own rules simply cannot work (to reiterate, it needs to break its own rules in order to work). Another way of saying this is that it needs "blind spots" in order to work.

However, Høgheim, following Michell, wants quantitative research in psychology and education to face its blind spots, which Michell (1997, p. 355) calls a "thought disorder" and thus be consistent with the empirical assumptions of the pure sciences. Høgheim therefore advocates an approach to measurement in educational quantitative research that does not break with the principles of classical measurement, namely additive conjoint measurement (ACM) and, exactly as we would expect, it seems that ACM is not workable. The reason that it does not work, according to Sijtsma (2012, as cited in Høgheim, 2023), is because the psychological context is too complex and the assumptions of ACM are impossible to meet in the real, open-system world, which one cannot scientifically close (as one can do in a laboratory, which is the only place where, as already explained, Popperian post-positivism can function without obvious problems). I argue that the reason provided by Sijtsma for why the ACM does not work – namely that it cannot cope with complexity – is exactly what we would expect if a researcher were to embark on a strict, non-hypocritical, research project based on empirical realism. That is, it is because researchers violate their own theory in their practice that they manage to achieve any knowledge at all. If they are held to their own rules, as is the case with ACM, meaningful research cannot happen.

### **The epistemic fallacy results in instrumentalism**

To explain further, the epistemic fallacy insists that researchers must take any emergent being, whether it is a person, a social system, an ecosystem or the climate, and try to reduce it to its measurements (such researchers have no ontology for emergent beings – they think that they are not really real). Another way of saying this is to say that the epistemic fallacy results in instrumentalism. However, there is much more to any entity than its instrumental measurements. Underneath every measurable entity there exists the underlying structures and mechanisms of reality that were triggered and maintained long enough for that being to come into existence. Whether we observe a tree in a forest or a child in a classroom, we can do so in an instrumental or a non-instrumental way. The instrumental way separates the entities from all other entities, ignoring the socio-ecological structures and mechanisms that enabled their existence in the first place. The non-instrumental way allows for that which is measurable to be just the tip of the iceberg of reality, and therefore it acknowledges that there is a large amount of reality that is not directly measurable. In this way, non-instrumentalism, by definition, shows us that all things are connected, since both the observable (and thus measurable) tree and the observable (and thus measurable) child depend on the existence of implied (but not directly observable) relationships, structures and mechanisms, such as the social system and the ecological cycles of life and death.

In terms of ecology (but this is also true of society), we can see how these ecological cycles connect the individual entities in a system: the child breathes in the oxygen created by the tree’s photosynthesis, while the tree absorbs the carbon dioxide exhaled by the child and converts it into carbohydrates, which the child might consume later should they eat the fruit of the tree. Eventually, the body of the child will reach maturity, grow old and die, and its decay will nurture the trees. To think non-instrumentally, that is, transcendently, about the child is not only to think “larger” than the child, towards its family, its school and wider society/ecosystems, but also to think “smaller” than the child, to the details of its immediate relationship with its surroundings and, importantly, to its mind and mental health. Thus, this non-instrumental view immediately points to connectedness, not only of mind to body but of mind to other minds and other bodies. As we would expect, Høgheim’s ACM fails to account for this complexity and thus it is unworkable, where, by the word “unworkable,” I mean that it is not fit for the purpose of finding out meaningful things about the complex world. In order to account for complexity and thus avoid instrumentalism, we need theories about the underlying structures and mechanisms of reality that underlie the empirical aspects of reality.

Ironically, the mainstream approach to psychological/educational measurement hypocritically and illicitly – in terms of its own rules about what counts as science – actually does allow complexity into the research process and thus it does sometimes “work,” despite its inadequate philosophy of science. It manages to do this by the

use of theories,<sup>5</sup> and it is these theories that Høgheim questions because of his pure, Popperian empiricist stance on the matter. Høgheim is therefore in line with Michell (2008, p. 10), who suggests that psychometricians claim to know something that they do not know, whilst erecting barriers that protect ignorance. He calls this “pathological science” (p. 10). If Høgheim manages to persuade psychologists and educators to “cease and desist” (Meier, 1994, cited in Michell, 1997, p. 354) from coming up with their theories, and/or stop them from assuming that their theories are about real things, in my opinion, he will halt the scientific achievement of the disciplines. This is because, from a Bhaskarian critical realist perspective, we can only know about the deeper, nonempirical (unobservable) levels of reality via our theories about them. Therefore, science may be more certain, but it will be largely irrelevant, since it is these deeper levels of emergent reality that form the true subject matter of what is important to us and what motivates our research. In this dystopian scenario, educators and psychologists might measure things – such as they might do with ACM – but without theories about the deeper levels of reality, their measurements would be meaningless and unable to contribute to the main function of these disciplines, which is to guide human praxis.

As a trained ecologist as well as an educator, it is easy for me to conceive of this dystopian scenario because, perhaps arguably, it already exists in the intermediate discipline of ecology. Any perusal of mainstream, respected ecology journals will show that they rarely deal with issues of importance, such as the survival of humanity, because they insist on a kind of science that too strictly adheres to the principles of mainstream “science.” That is, the very discipline that should be leading our fight for survival against such threats as climate change tends to be reticent on the subject. Charles Hall (cited in Price, 2019, p. 353) explains that this is because ecology journals insist on standards of excellence that are “too narrowly conceived” in terms of experimental science. He explains that one of the greatest ecologists, Charles Darwin, did not use such methods but instead added to humanity’s knowledge about species and ecology without practising science the way it was “supposed” to be done according to the scientific leaders of his day.

Essentially, it is only because psychologists and educators typically violate the principles of positivist science that they manage to find out anything of significance. Therefore, the solution to their “hypocrisy” or, as Michell puts it, their “pathology,” is not to insist that they become more scientific, but that they change what they consider to be science. That is, scientists need to change their theory to be more in line with their practice, whereas Høgheim wants scientists to change their practice to be more in line with their theory.

---

<sup>5</sup>The theories that I discuss here are not the same as the shallow correlational theories of cause and effect assumed to be the only theories permissible in empiricist science.

However, although I disagree with Høgheim in that I am not in favour of basing educational research on Euclidean measurement, nevertheless I agree with him that theories, or what he calls “constructs,” need to be treated with care. All too often, researchers make incorrect assumptions about what their measurements mean, and this can result in dangerous relativism. By relativism I mean that without ontology there is no way to judge the truth of a construct, so that it is simply at the whim of a researcher, whose only limitation is that the construct somehow saves appearances, that is, it somehow seems feasible given the measurements. For instance, when psychologists and educators think they are measuring intelligence through IQ tests, they may simply be measuring cultural understandings of the world; when educators think that they are measuring students’ mastery of a subject, they may merely be measuring their ability to succeed in examinations (Koretz, 2008) and, as I have argued above and elsewhere (Price, 2014), when educators think that measuring absence from school is measuring a cause of school failure, they may in fact only be measuring poverty, that is, it is poverty, not absenteeism *per se*, that is the main reason behind school failure. These errors are not innocent but can have serious consequences. IQ tests can result in disadvantaging people from marginal cultures; examinations can disadvantage poor students whose state schools do not “teach to the exam;” and trying to stop absenteeism by heavily fining parents, as happens in the United Kingdom, merely exacerbates the root problem of poverty.

The solution lies in being more transparent about the theorising that currently goes on in a somewhat hidden way in the intermediate sciences. It has to be somewhat hidden because of the contradiction that in empiricist theory, constructs (or theories about what is happening) are, to an extent, frowned upon and are not supposed to be about anything real. As such, they, ostensibly, should not be taken seriously – indeed it is the fact that they are, in practice, taken seriously and assumed to be about real things that Høgheim finds contentious. This is where Høgheim and I agree: we both argue against hypocrisy. However, Høgheim’s solution remains instrumentalist because it lacks an ontology for the emergent things in psychology and education that are unobservable. In the absence of ontology, there is no formal requirement for scientists to transparently put forward all the competing theories that would explain certain measurements and decide which of these theories are the most plausible by a process of elimination, which critical realists call judgemental rationalism. If scientists were formally required to do this, we would have the solution to construct validity. Whilst this is a fallible approach, because it assumes that we may have to change our theory given further evidence, it is nevertheless the only way that knowledge acquisition has ever happened, and indeed it is the only way that it can happen (Bhaskar, 2016, p. 25).

For example, we would see that a theory that posits that *there are cultural biases in IQ tests* would explain more of the evidence than a theory that posits that intelligence is an innate characteristic directly measured by IQ tests, and we would see that a theory *that examinations are neither necessary nor sufficient as a measure of student subject*

*mastery* would explain the school test scores and other evidence better than a theory that exams are always the best and only reliable way to measure subject mastery. Currently, since there is no formal requirement for scientists to use judgemental rationalism to choose between competing theories that explain their empirical data, policymakers can choose any theory from those available. I have argued elsewhere that they tend to choose the theories that best suit their interests (Price, 2014). Thus, assuming that IQ is a measure of innate intelligence suits a racist agenda, and assuming that school exam grades are a measure of subject mastery suits the agenda of those who can pay for their children to attend schools that are good at teaching the skill of passing an exam. Irrealism about transcendent entities results in relativism about knowledge about those transcendent entities and this, as Michel Foucault (1975, 1976) explains, makes knowledge about them simply a function of those who are most powerful.

Finally, a brief comment on Høgheim's use of critical realism in his article. When he challenges psychologists' assumption that the transcendental things that they are researching are real, he is also challenging Bhaskar's critical realism, since it has a similar assumption. Bhaskar is quoted by Høgheim as follows, "To be is *not* to be the value of a variable; though it is plausible (if, I would argue incorrect) to suppose that things can only be *known* as such" (Bhaskar, 2008, p. 29). In terms of this statement, we can say that Høgheim, Bhaskar and I all agree that *to be is not to be the value of a variable*, but Høgheim, it seems, thinks it is plausible to suppose that, nevertheless, *things can only be known as variables*, whereas Bhaskar (and I) think that this is incorrect and that *there is more to a thing than its variables*, which we can know about through retroduction.

## Conclusion

What makes Høgheim's article paper interesting for me is that both the author and I are, I believe, motivated by the same thing. That is, we are both motivated to end hypocrisy in education and psychology. However, we approach this similar task in different ways. Høgheim suggests that we can circumvent the problem by avoiding representationalism in educational or psychological measurement and returning to the classic definition of measurement, but I think that avoiding representation (coding, modelling, theorising) is impossible. Instead of trying to avoid it, I suggest that a way to achieve a version of representation that does not result in a lack of ontology (which is the problem for Høgheim) is to use the semiotic triangle, which assumes that not only does representation happen, but that it must also be meaningfully representative of the thing it is trying to represent. Høgheim suggests that educational theories are not about anything, but I think that they measure a deeper, emergent reality that cannot be measured empirically and that, far from trying to avoid theorising, we need to be transparent about the theorising process (which includes both retroduction and judgemental rationalism), so that we can ensure that we choose theories that best fit

the evidence. In so doing, I argue that it is possible to avoid making errors related to construct validity and to challenge more easily the theories favoured by those with questionable interests.

Høgheim and his main inspiration, Michell, are being honourable in trying to remain faithful to their scientific values. However, to remain absolutely faithful to the classic version of measurement is to make some kinds of measurement impossible, especially in intermediate disciplines such as psychology or education, where the nature of that which is of interest makes it impossible for the research to be carried out in laboratories. Instead, I suggest that rather than trying to change scientists’ practice to be faithful to their positivist theory of science (as suggested by Høgheim), it is better to change the theory of science to a more adequate version (such as critical realism), that better reflects scientists’ actual practice.

### Author biography

**Leigh Price** is Associate Professor at Inland Norway University of Applied Sciences and General Editor of the *Journal of Critical Realism*. She is Honorary Senior Research Associate at Rhodes University and Honorary Emeritus at the Institute of Education, University College London. She has a Ph.D. in Education and an M.Sc. in Tropical Resource Ecology.

### References

- Bhaskar, R. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy*. Routledge.
- Bhaskar, R. (2008). *Dialectic: The pulse of freedom*. Routledge. (Original work published in 1993)
- Bhaskar, R. (2009). *Scientific realism and human emancipation*. Routledge. (Original work published in 1986)
- Bhaskar, R. (2014). *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. Routledge. (Original work published in 1979)
- Bhaskar, R. (2016). *Enlightened common sense: The philosophy of critical realism*. Routledge.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061.
- Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.
- Campbell, N. R. (1920). *Physics, the elements*. Cambridge University Press.
- Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. Longmans Green and Company.
- Chapman, B., Fiscella, K., Duberstein, P., Kawachi, I., & Muennig, P. (2014). Measurement confounding affects the extent to which verbal IQ explains social gradients in mortality. *Journal of Epidemiology and Community Health*, 68(8), 728–733.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Ferguson, A., Myers, C. S., Bartlett, R. J., Banister, H., Bartlett, F. C., ... Brown, W. (1940). Quantitative estimates of sensory events. *Advances in Science*, 1, 331–349.
- Filazzola, A., & Cahill Jr, J. F. (2021). Replication in field ecology: Identifying challenges and proposing solutions. *Methods in Ecology and Evolution*, 12(10), 1780–1792.
- Foucault, M. (1975). *Discipline and punish* (A. Sheridan, Trans.). Gallimard.
- Foucault, M. (1976). *The history of sexuality (The will to knowledge: History of sexuality vol. 1)*. Allen Lane.
- Hamilton, A. J. (2018). *Scale and the Incas*. Princeton University Press.



- Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge University Press.
- Høgheim, S. (2023). Måling i kvantitativ utdanningsforskning: Et instrumentelt mistak? *Nordisk tidsskrift for pedagogikk og kritikk*, 9.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.
- McGrane, J. A. (2015). Stevens' forgotten crossroads: The divergent measurement traditions in the physical and psychological sciences from the mid-twentieth century. *Frontiers in Psychology*, 6, 431.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Michell, J. (1993). The origins of the representational theory of measurement: Helmholtz, Hölder, and Russell. *Studies in History and Philosophy of Science Part A*, 24(2), 185–206.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–383.
- Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist*, 36(3), 211–218. <http://doi.org/10.1080/00050060108259657>
- Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, 6(1–2), 7–24. <http://doi.org/10.1080/15366360802035489>
- Morris, M., & Rutt, S. (2004). *Analysis of pupil attendance data in Excellence in Cities (EiC) areas: An interim report*. DfES Publications.
- Nordahl-Hansen, A., & Kvernbekk, T. (2020). Construct validity in scientific representation: A philosophical tour. *Nordisk tidsskrift for pedagogikk og kritikk*, 6. <https://doi.org/10.23865/ntpk.v6.1704>
- Pilgrim, D. (2013). The failure of diagnostic psychiatry and some prospects of scientific progress offered by critical realism. *Journal of Critical Realism*, 12(3), 336–358.
- Popper, K. (1961). *The logic of scientific discovery*. Science Editions, Inc. (Original work published in 1959)
- Popper, K. (1967). *Conjectures and refutations: The growth of scientific knowledge*. Basic Books.
- Price, L. (2014). Hume's two causalities and social policy: Moon rocks, transfactuality, and the UK's policy on school absenteeism. *Journal of Critical Realism*, 13(4), 385–398.
- Price, L. (2019). A return to common-sense: Why ecology needs transcendental realism. *Journal of Critical Realism*, 18(1), 31–44.
- Stodola, J. (1971). *Mathematical contributions of the Mayas, Aztecs & Incas: A Native American curriculum unit for middle and high school*. NATAM XIX.
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39(4), 202.
- William, D. (2022). How should educational research respond to the replication “crisis” in the social sciences? Reflections on the papers in the special issue. *Educational Research and Evaluation*, 27(1–2), 208–214.
- Wittgenstein, L. (2010). *Philosophical investigations* (G. Anscombe, Trans.). John Wiley & Sons. (Original work published in 1953)